Universidad de Alcalá
COMISIÓN DE ESTUDIOS OFICIALES
DE POSGRADO Y DOCTORADO

ACTA DE EVALUACIÓN DE LA TESIS DOCTORAL

Año académico 2016/17

DOCTORANDO: **WIJERS, IRENE GEURTJE MARIA**
D.N.I./PASAPORTE: ******7224 Z**

PROGRAMA DE DOCTORADO: **D325 DOCTORADO EN CIENCIAS DE LA SALUD**
DEPARTAMENTO DE: **CIRUGÍA, CIENCIAS MÉDICAS Y SOCIALES**
TITULACIÓN DE DOCTOR EN: **DOCTOR/A POR LA UNIVERSIDAD DE ALCALÁ**

En el día de hoy 04/07/17, reunido el tribunal de evaluación nombrado por la Comisión de Estudios Oficiales de Posgrado y Doctorado de la Universidad y constituido por los miembros que suscriben la presente Acta, el aspirante defendió su Tesis Doctoral, elaborada bajo la dirección de **MARÍA JOÃO FORJAZ // MANUEL FRANCO TEJERO.**

Sobre el siguiente tema: *THE DISEASE BURDEN MORBIDITY ASSESSMENT: A VALIDATION STUDY*

Finalizada la defensa y discusión de la tesis, el tribunal acordó otorgar la CALIFICACIÓN GLOBAL[11] de (**no apto, aprobado, notable y sobresaliente**): SOBRESALIENTE

Alcalá de Henares, 4 de JULIO de 2017

EL PRESIDENTE — Fdo.:

EL SECRETARIO — Fdo.: Carmen Rodríguez Blázquez

EL VOCAL — Fdo.: Gloria Fernández-Mayoralas

Con fecha 24 de julio de 2017 la Comisión Delegada de la Comisión de Estudios Oficiales de Posgrado, a la vista de los votos emitidos de manera anónima por el tribunal que ha juzgado la tesis, resuelve:

- [X] Conceder la Mención de "Cum Laude"
- [ ] No conceder la Mención de "Cum Laude"

La Secretaria de la Comisión Delegada

FIRMA DEL ALUMNO,

Fdo.:

[11] La calificación podrá ser **"no apto" "aprobado" "notable" y "sobresaliente"**. El tribunal podrá otorgar la mención de "cum laude" si la calificación global es de sobresaliente y se emite en tal sentido el voto secreto positivo por unanimidad.

UNIVERSIDAD DE ALCALÁ, PATRIMONIO DE LA HUMANIDAD

INCIDENCIAS / OBSERVACIONES:

La Dra. Gloria Fernández - Mayoralas Fernández ha actuado como Vocal en ausencia del Dr. Vicente Rodríguez Rodríguez.

Universidad de Alcalá
COMISIÓN DE ESTUDIOS OFICIALES
DE POSGRADO Y DOCTORADO

En aplicación del art. 14.7 del RD. 99/2011 y el art. 14 del Reglamento de Elaboración, Autorización y Defensa de la Tesis Doctoral, la Comisión Delegada de la Comisión de Estudios Oficiales de Posgrado y Doctorado, en sesión pública de fecha 24 de julio, procedió al escrutinio de los votos emitidos por los miembros del tribunal de la tesis defendida por *WIJERS, IRENE GEURTJE MARIA*, el día 4 de julio de 2017, titulada *THE DISEASE BURDEN MORBIDITY ASSESSMENT: A VALIDATION STUDY*, para determinar, si a la misma, se le concede la mención "cum laude", arrojando como resultado el voto favorable de todos los miembros del tribunal.

Por lo tanto, la Comisión de Estudios Oficiales de Posgrado **resuelve otorgar** a dicha tesis la

MENCIÓN "CUM LAUDE"

Alcalá de Henares, 27 julio de 2017
EL PRESIDENTE DE LA COMISIÓN DE ESTUDIOS
OFICIALES DE POSGRADO Y DOCTORADO

Firmado digitalmente por VELASCO PEREZ JUAN RAMON - DNI 03087239H
Fecha: 2017.07.30 19:02:22 +02'00'

Juan Ramón Velasco Pérez

Copia por e-mail a:
Doctorando: WIJERS, IRENE GEURTJE MARIA
Secretario del Tribunal: CARMEN RODRÍGUEZ BLÁZQUEZ.
Directores de Tesis: MARÍA JOÂO FORJAZ // MANUEL FRANCO TEJERO

UNIVERSIDAD DE ALCALÁ. PATRIMONIO DE LA HUMANIDAD

Universidad de Alcalá

ESCUELA DE DOCTORADO
Servicio de Estudios Oficiales de Posgrado

DILIGENCIA DE DEPÓSITO DE TESIS.

Comprobado que el expediente académico de D./Dª IRENE GEURTJE MARIA WIJERS reúne los requisitos exigidos para la presentación de la Tesis, de acuerdo a la normativa vigente, y habiendo presentado la misma en formato: ☒ soporte electrónico ☐ impreso en papel, para el depósito de la misma, en el Servicio de Estudios Oficiales de Posgrado, con el nº de páginas: 144 se procede, con fecha de hoy a registrar el depósito de la tesis.

Alcalá de Henares a 27 de abril de 2017

Emilia Rivero Antúnez

Fdo. El Funcionario

PROGRAMA DE DOCTORADO EN CIENCIAS DE LA SALUD

THE DISEASE BURDEN MORBIDITY ASSESSMENT: A VALIDATION STUDY

Tesis doctoral presentada por:

Irene Geurtje Maria Wijers

Directora:

Dra. Maria João Forjaz, Instituto de Salud Carlos III

Co-director:

Dr. Manuel Franco Tejero, Universidad de Alcalá

Alcalá de Henares, 2017

Universidad de Alcalá

FACULTAD DE MEDICINA Y CIENCIAS DE LA SALUD
DEPARTAMENTO DE CIRUGÍA, CIENCIAS MÉDICAS Y SOCIALES
Campus Universitario
Ctra. Madrid-Barcelona, Km. 33,600
E-28871 Alcalá de Henares (Madrid)
Tel. 91 885 4881-4539
Fax: 91 885 4714
E-mail: dpto.cirumedsoc@uah.es

UNIVERSIDAD DE ALCALÁ. PATRIMONIO DE LA HUMANIDAD

Dña. MARIA JOÃO FORJAZ, Científica Titular OPIs en la Escuela Nacional de Sanidad del Instituto de Salud Carlos III
Y
D. MANUEL FRANCO TEJERO, Profesor Titular Interino del Departamento de Cirugía, Ciencias Médicas y Sociales de la Universidad de Alcalá

Como Directores de la presente Tesis Doctoral,

CERTIFICAN

Que el trabajo titulado "THE DISEASE BURDEN MORBIDITY ASSESSMENT: A VALIDATION STUDY" realizado por **Dña. Irene Geurtje María Wijers**, reúne los requisitos metodológicos y valor científico adecuados para ser presentado y defendido para optar al grado de Doctor por la Universidad de Alcalá.

Y para que así conste, se expide el presente certificado en Alcalá de Henares, a veinticuatro de abril de dos mil diecisiete.

Universidad de Alcalá

FACULTAD DE MEDICINA Y CIENCIAS DE LA SALUD
DEPARTAMENTO DE CIRUGÍA, CIENCIAS MÉDICAS Y SOCIALES
Campus Universitario
Ctra. Madrid-Barcelona, Km. 33,600
E-28871 Alcalá de Henares (Madrid)
Tel. 91 885 4881-4539
Fax: 91 885 4714
E-mail: dpto.cirumedsoc@uah.es

UNIVERSIDAD DE ALCALÁ, PATRIMONIO DE LA HUMANIDAD

D. FERNANDO NOGUERALES FRAGUAS, Catedrático de Cirugía y Director del Departamento de Cirugía, Ciencias Médicas y Sociales de la Universidad de Alcalá

CERTIFICA

Que el trabajo presentado por **Dña. Irene Geurtje María Wijers** titulado "The Disease Burden Morbidity Assessment: a validation study" ha sido realizado en el Departamento de Cirugía, Ciencias Médicas y Sociales y reúne los requisitos científicos de originalidad y rigor metodológicos suficientes para poder ser presentado como tesis doctoral ante el tribunal que corresponda.

Y para que así conste, se expide el presente certificado en Alcalá de Henares, a veinticuatro de abril de dos mil diecisiete.

ACKNOWLEDGEMENTS

This project would not have been possible without the support of many people. Many, many thanks to João Forjaz, who taught me how to write articles, revised my work over and over again and corrected every tiny little detail. Who always was optimistic when I was pessimistic and always supported me in every decision I took. And who even 'mentored' me when I became a mum for the first and second time during this project, with her broad experience as a mum of two. Many thanks also to Manuel Franco, for being my co-director and revising my thesis.

Special thanks also to Alba Ayala, the best statistician I know, who conducted the Rasch analysis and was always there for me when I had questions. Thanks to Ángel Rodríguez and Vicente Rodríguez for giving me the opportunity to work with the ELES-PS database.

Finally, thanks to my husband, who has been cooking, cleaning and taking our son to the park without complaining, just to let me work on my thesis. And very special thanks to Nina, 'the little secretary', who spent the first four months of her life on my chest, in the baby carrier, while I was finishing this work.

RESUMEN

INTRODUCCIÓN

El envejecimiento de la población es un proceso que está teniendo lugar en todo el mundo. Una de las consecuencias es el aumento de la prevalencia de los problemas crónicos de salud y, por lo tanto, también la coexistencia de ellas, denominada comorbilidad o multimorbilidad. La multimorbilidad es un problema de salud muy frecuente en todos los grupos de edad, y especialmente en las personas mayores. Es un factor pronóstico importante, con efectos negativos sobre la mortalidad, los resultados quirúrgicos, las complicaciones postoperatorias y la duración de la estancia hospitalaria, y un efecto directo e independiente sobre la discapacidad y la calidad de vida. Existen diferentes instrumentos para evaluar la multimorbilidad; la elección del instrumento depende del contexto del estudio y los resultados de interés. La escala Disease Burden Morbidity Assessment (DBMA) es un cuestionario de autovaloración en el que los participantes clasifican la carga de la enfermedad causada por una serie de problemas crónicos de salud. Fue diseñada y validada para asociarse con resultados centrados en el paciente. Sin embargo, todavía no se había realizado una validación siguiendo una metodología psicométrica o clinimétrica.

OBJECTIVOS

Los objetivos de esta tesis fueron validar la DBMA de acuerdo con la Teoría Clásica de los Test (TCT) (Estudio 1), evaluar la validez de grupos conocidos, la validez convergente y la validez predictiva (Estudio 2) y realizar un análisis Rasch de la escala (Estudio 3).

MÉTODOS

Se utilizaron datos del Estudio Longitudinal Envejecer en España, Estudio Piloto (ELES-PS), en la que se incluyeron adultos no institucionalizados mayores de 50 años residentes en España. En el Estudio 1 y 2, se utilizaron submuestras de personas de 65 años o más. El cuestionario CAPI del estudio ELES-PS incluyó la DBMA. En esta escala, que consiste en una lista de 21 problemas crónicos de salud, se pregunta a los participantes para cada problema crónico de salud si lo tienen y, en caso afirmativo, hasta qué punto interfiere con su vida cotidiana en una escala de 1 (nada) a 5 (mucho). La puntuación total, obtenida sumando las puntuaciones de todos los problemas crónicos de salud presentes, proporciona una medida de la carga de la enfermedad autopercibida.

En el primer estudio, se analizaron las propiedades psicométricas de la escala (viabilidad, aceptabilidad, asunciones escalares, fiabilidad y validez de constructo). La dimensionalidad fue estudiada con un análisis factorial exploratorio. En el Estudio 2, se evaluó la validez de grupos conocidos para sexo y grupos de edad (<75 años frente a ≥75 años). Para la validez convergente, se utilizó un modelo de regresión lineal multivariante para evaluar las diferencias de los resultados de la DBMA en función de la edad y el sexo, la salud autopercibida, la función física, la calidad de vida, el equilibrio afectivo y el uso de recursos sanitarios. Para la validez predictiva se estudió la asociación con la mortalidad a cuatro años utilizando un modelo de Cox y curvas de Kaplan-Meier. En el análisis Rasch se analizaron los siguientes atributos métricos de una manera iterativa: ajuste al modelo Rasch, unidimensionalidad, fiabilidad, funcionamiento diferencial de los ítems (DIF), independencia local de los ítems y adecuación de la escala de respuesta. Posteriormente, se evaluó la validez de constructo de la medida lineal proporcionada por el análisis Rasch.

RESULTADOS

En el análisis TCT, la viabilidad y la aceptabilidad fueron satisfactorias, salvo efectos suelo grandes (> 50%) y una distribución asimética (*skewness*=1,8). La correlación ítem-total corregida osciló entre 0,10-0,49, el índice de homogeneidad de los ítems fue de 0,09 y el alfa de Cronbach de 0,72. La DBMA mostró una correlación fuerte con la escala de función física (r = -0,56) y la salud autopercibida (r = -0,56), y moderada con la depresión (r = 0,41) y la calidad de vida (r = -0,41). El análisis factorial exploratorio extrajo 5 factores, explicando el 44% de la varianza.

En el análisis de grupos conocidos en el Estudio 2 se encontraron para las mujeres prevalencias de enfermedades mayores y también mayor carga de enfermedad por enfermedad presente. Las mismas diferencias fueron halladas para los grupos de edad, pero fueron menos pronunciadas. En la regresión multivariante, el sexo, la salud autopercibida, la función física, la calidad de vida, el equilibrio afectivo y la utilización recursos sanitarios se asociaron significativamente con la DBMA. El modelo de Cox mostró un *hazard ratio* de 1,07 y las curvas de Kaplan-Meier mostraron tasas de supervivencia más bajas en los participantes con mayores puntuaciones en la DBMA.

En el análisis Rasch, se recodificaron las escalas de respuesta para lograr un ajuste adecuado al modelo de Rasch. La fiabilidad (*person separation index*) fue baja. La escala mostró unidimensionalidad y no se encontró dependencia local de respuestas ni DIF relevante. El análisis de precisión relativa mostró que la medida lineal discriminaba mejor entre los grupos de edad que la puntuación original, pero para el sexo no se encontró ninguna diferencia.

CONCLUSIONES

A pesar de algunas limitaciones como la fiabilidad por debajo de los esperado y un efecto suelo alto, se encontró apoyo para la validez de la DBMA. Es un cuestionario autoinformado que repite la misma pregunta para diferentes problemas crónicos de salud, lo que la hace particularmente aplicable en las personas mayores, ya que es fácil de entender y se puede rellenar en un corto período de tiempo. En nuestra sociedad envejecida, con un número de personas mayores con multimorbilidad cada vez mayor, la DBMA puede ser una herramienta útil que ayuda a comprender mejor y mejorar la atención de las personas mayores con múltiples enfermedades crónicas.

SUMMARY

INTRODUCTION

Population aging is a process that is taking place all over the world. One of the consequences is the increase in the prevalence of chronic conditions, and therefore also the co-existence of them, so-called comorbidity or multimorbidity. Multimorbidity is a highly prevalent health problem among all age groups, and especially in the elderly. It is an important prognostic factor, with well-described negative effects on mortality, surgical outcome, postoperative complications, and hospital length of stay, and a direct and independent effect on disability and quality of life. Different instruments exist to assess multimorbidity, and the choice of instrument depends on the study context and outcomes of interest. The Disease Burden Morbidity Assessment (DBMA) is a self-report questionnaire in which participants rate the disease burden caused by a number of medical conditions. It was designed and validated to be associated with patient-centered outcomes. However, a validation following psychometric or clinimetric methodology had not been performed yet.

OBJECTIVES

The objectives of this thesis were to validate the DBMA according to the Classical Test Theory(CTT) (Study 1), to assess known-groups, convergent and predictive validity (Study 2) and to perform a Rasch analysis of the scale (Study 3).

METHODS

Data were used from the Ageing in Spain Longitudinal Study, Pilot Survey (ELES-PS), which included community-dwelling adults aged 50 years or more living in Spain. In Study 1 and 2, subsamples of persons aged 65 years and older were used. The CAPI questionnaire of the ELES-PS included the DBMA. In this scale, consisting of a list of 21 chronic medical conditions, participants are asked for every condition whether they have it and if so, to what extent it interferes with their everyday life on a scale from 1 (not at all) tot 5 (a lot). The total score, obtained by summing the scores given to the different conditions, provides a measure of self-reported disease burden.

In the first study, psychometric properties of the scale (feasibility, acceptability, scaling assumptions, reliability and construct validity) were analyzed. Dimensionality was assessed through an exploratory factor analysis. In Study 2, known-groups validity for sex and age groups (< 75 years vs. ≥75 years) was assessed. For convergent validity, a multivariate linear regression model was used to evaluate differences in DBMA scores as a function of age and sex, perceived health, physical functioning, quality of life, affect balance and utilization outcomes. For predictive validity, the association with four-year mortality was assessed using a Cox model and Kaplan-Meier curves. In the Rasch analysis, test of fit to the Rasch model, reliability, unidimensionality, response dependency, category structure, scale targeting and differential item functioning (DIF) were studied in an iterative way. Construct validity of the linear measure provided by the Rasch analysis was subsequently assessed.

RESULTS

In the CTT analysis, satisfactory feasibility and acceptability were found, except for large floor effects (>50%) as well as a skewed distribution (skewness=1.8). Item-total corrected correlation ranged 0.10-0.49, item homogeneity index was 0.09, and Cronbach's alpha was 0.72. Disease burden correlated strongly with physical functioning (r= -0.56) and perceived health (r=-0.56), and moderately with depression (r= 0.41) and quality of life (r=-0.41). Exploratory factor analysis extracted 5 factors, explaining 44% of the variance.

The known-groups analysis in Study 2 found higher disease prevalences and also higher disease burden per present condition for women. The same differences were found for age groups but less pronounced. In the multivariate regression, sex, perceived health, physical functioning, quality of life, affect balance and primary/outpatient care utilization were significantly associated with the DBMA. The Cox model displayed a hazard ratio of 1.07 and the Kaplan-Meier curves showed lower survival rates in participants with higher DBMA scores.

In the Rasch analysis, items needed to be rescored by collapsing response categories to achieve an adequate fit to the Rasch model. Reliability (person separation index) was low. The scale was unidimensional and neither response dependency nor relevant DIF were found. Relative precision analysis showed that the linear measure discriminated better between age groups than the original raw score, but for sex no difference was found.

CONCLUSIONS

Despite some limitations such as reliability below the expected and high floor effects, support was found for the validity of the DBMA. It is a self-reported questionnaire that repeats the same question for different conditions, which makes it particularly applicable in older populations, since it is easy to understand and can be filled out in a short amount of time. In our ageing society, with increasing numbers of older people with multimorbidity, the DBMA can be applied to better understand and improve care for older persons with multiple chronic conditions.

LIST OF ABBREVIATIONS

AUC	Area under the curve
CAPI	Computer-assisted personal interviewing
CCI	Charlson Comorbidity Index
CDS	Chronic Disease Score
CES-D	Center for Epidemiologic Studies Depression Scale
CI	Confidence interval
CIE	Change-in-estimate
CIRS	Cumulative Illness Rating Scale
CIRS-G	Cumulative Illness Rating Scale geriatrics
COPD	Chronic obstructive pulmonary disease
CTT	Classical Test Theory
DBMA	Disease Burden Morbidity Assessment
DBMA-Fv	Disease Burden Morbidity Assessment French version
DIF	Differential item functioning
ECM	Elixhauser´s Comorbidity Measure
ELES-PS	Estudio Longitudinal Envejecer en España, pilot study
HR	Hazard ratio
ICC	Intraclass correlation coefficient
ICD	International Classification of Diseases
IRT	Item Response Theory
ITCC	Item-total corrected correlation
LOS	Length of stay
OR	Odds ratio
OSA	Obstructive sleep apnea
PCA	Principal component analysis

PSI	Person separation index
PWI	Personal Wellbeing Index
QoL	Quality of life
ROC	Receiver operating characteristic
SCQ	Self-Administered Comorbidity Questionnaire
SD	Standard deviation
SE	Standard error
SPANE	Scale of Positive and Negative Experience

CONTENTS

TABLES AND FIGURES

TABLES

FIGURES

1. INTRODUCTION

1.1. POPULATION AGEING

The ageing of populations is a global phenomenon and is taking place in almost all countries of the world. This process results from, on the one hand, declining birth rates and, on the other, decreasing mortality. This leads to a reduction in the proportion of children and an increase of older people in the population. In 1950, the worldwide proportion of older persons (60 years and older) was 8 % (1). This proportion rose to almost 12% in the year 2013, and is expected to reach 21% in the year 2050. In developed countries, these proportions are even higher, with a proportion of almost 23% in the year 2013 and a predicted proportion of 32% for the year 2050.

Population aging has important social and economic consequences. The old-age support ratios (number of working-age adults per older person in the population) are already low in developed countries, and are expected to continue to decrease in the coming decades with fiscal pressures on support systems for older persons. Another important consequence lies in the prevalence increase of non-communicable diseases and disability as populations age (2). However, the aging of populations also has its positive sides. The increased prevalence of non-communicable diseases originates in the positive trend of drastically reduced adult mortality. Older persons can increasingly live independently until higher ages, and can support themselves economically, even making financial contributions to younger family members (1).

1.2. CO- AND MULTIMORBIDITY

Not only the prevalence of non-communicable diseases, but also the co-existence of conditions, so-called comorbidity or multimorbidity, increases with the process of population ageing. Comorbidity can be described as the existence of one of more other diseases among persons with one index-disease (3). This definition implies that the main interest is on an index condition and the possible effects of other disorders on the prognosis of this condition (4). The term multimorbidity also refers to the presence of multiple chronic conditions in one person, but without the perspective of a specific index-disease. Nevertheless, both terms are often used as synonyms, and since the DBMA can be applied from both perspectives, we applied both in this study.

Multimorbidity is a highly prevalent health problem among all age groups, and especially in the elderly (5–7). In a systematic review, published by Marengoni et al., the prevalence in older people ranged from 55 to 98%, and a prevalence of 20-30% was found when the whole population was considered (4). However, several problems arise when comparing the prevalences found in different studies. Studies included different populations, such as primary care or general population, and different methods of data collection were used: self-report, medical records or pharmacy database utilization, for example. Also, although most studies define the presence of two concurrent chronic diseases as multimorbidity, some studies use other definitions, such as the presence of three concurrent conditions. Another limiting factor is the number of disorders taken into account when counting the number of conditions. When using a larger list of included conditions, studies are more likely to find higher prevalences than when using less extended lists. These differences may lead to the broad intervals found in systematic reviews about the prevalence of multimorbidity (4,8).

1.3. MULTIMORBIDITY AS A PROGNOSTIC FACTOR

Multimorbidity is an important prognostic factor. Is has well-described negative effects on mortality, surgical outcome, postoperative complications, and hospital length of stay (LOS) (9), and a direct and independent effect on disability and quality of life (QoL) (10).

1.3.1. MULTIMORBIDITY AND MORTALITY

In a recent systematic review about the relation between multimorbidity and mortality in older adults, an overall hazard ratio (HR) of 1.44 was found (11). Also, the number of conditions was positively related to mortality, with a HR of 1.20. When comparing to individuals without multimorbidity, the risk of death was 1.73 times higher in patients with two or more and 2.72 times higher in patients with three or more chronic conditions. However, the included studies were very heterogenic and, again, emphasis was made on the importance of multimorbidity measurement standardization: in order to make studies comparable, a standard methodology should be followed.

1.3.2. MULTIMORBIDITY AND SURGICAL OUTCOMES

Comorbid patients are known to have poorer surgical outcomes than other persons. In a study in which the impact of comorbidity on surgical outcomes in laparoscopy-assisted distal gastrectomy was assessed (12), comorbidity was found to be related to local complications, systemic complications and hospital mortality. In multivariate analysis, an odds ratio (OR) of 1.79 was found for local complications and an OR of 2.89 for systemic complications. Another study, assessing the influence of comorbidity on surgical outcomes in older surgical cancer patients , concluded that patients with comorbidity had more

postoperative morbidity and mortality (13). Patients with comorbid conditions were more likely to suffer from postoperative infections, cardiac complications and postoperative death. Yet, Hewitt et al. (14) who conducted an observations study among older emergency general surgical patients, did not find any differences in surgical outcomes (LOS, readmissions and mortality) between comorbid and non-comorbid patients, and pointed out the importance of not excluding these patients from surgery. Indeed, in the previously cited article about surgical outcomes in older surgical cancer patients, comorbid participants showed lower resection rates than other patients, especially in advanced tumor stages. This leads to poorer survival rates and might be one of the factors that plays a role in the higher mortality rate among persons with multiple chronic conditions.

1.3.3. MULTIMORBIDITY AND HOSPITAL LENGTH OF STAY

Different studies show that patients with multiple comorbid conditions have more prolonged hospital stays. In a study among hospitalized patients with cancer and febrile neutropenia (15), the prevalence of LOS ≥ 10 days increased with the number of comorbid conditions: from 11.2% in patients with no comorbidities to 62.3% in patients with five or more comorbidities among patients with solid tumors, and from 17.0% to 80.6% in patients with lymphomas. Another study performed in patients with acute burn injury showed the same pattern: among patients with dementia, peptic ulcer disease and diabetes, there was a 60%, 53% and 26% increase in LOS, respectively.

1.3.4. MULTIMORBIDITY AND HEALTHCARE UTILIZATION COSTS

LOS is one of the features that determine healthcare utilization costs. However, it is expected that overall healthcare use increases in patients with multimorbidity, which was confirmed by a study conducted in a primary care population aged over 50 years in Ireland (16). The addition of a single chronic condition led to an associated increase in primary care consultations, hospital outpatient visits, hospital admissions and total health care costs. The latter increased from €760 for zero to €4096 for more than four comorbid conditions. A Swiss study performed in an elderly community-dwelling population drew a similar conclusion (17): total healthcare utilization costs were 5.5 times higher in persons with multiple chronic conditions. Each additional condition was associated with an increased cost of 33% and an increase of 3.2 primary care/specialist consultations per year. In total, the mean number of consultations per year was 4.4 in non-multimorbid persons and 15.7 in multimorbid patients.

1.3.5. MULTIMORBIDITY AND DISABILITY

Several studies have confirmed the association between multimorbidity and disability. Garin et al. (18) performed a study about the impact of multimorbidity on disability and QoL in a sample aged over 50 years in Spain. All included 11 chronic conditions, except hypertension, were statistically associated with disability, and a higher number of chronic conditions was associated with greater disability. Depression, anxiety and stroke were found to have the greatest impact on functional status. This finding was confirmed by two studies performed by Marventano et al. in community-dwelling and institutionalized persons aged 65 years and older in Spain, in which dementia and neuropsychiatric disorders were found to be most strongly associated with disability (19,20). St. John et al

(21) studied the relationship between multimorbidity, disability, and mortality in community-dwelling older adults in Canada. Among persons with no chronic conditions, 1.1% had moderate to severe impairment, versus 25.3% in persons with seven or more conditions.

1.3.6. MULTIMORBIDITY AND QUALITY OF LIFE

Multimorbidity has a negative effect on QoL. Fortin et al. (22) performed a systematic review about the relationship between multimorbidity and QoL in primary care. Thirty studies were included and all of them found an inverse association between multimorbidity and QoL. This association was especially clear between multimorbidity and physical domains of QoL. In the above cited study of Garin et al. (18), the same association was found for QoL as for disability: all included 11 chronic conditions, except hypertension, were statistically related to QoL and a higher number of chronic conditions was associated with lower QoL. Depression, anxiety and stroke had the highest negative effect on QoL. Forjaz et al. also found mental health disorders to have a great impact on QoL in a study in multimorbid persons aged 65 years and older in Spain (OR: 1.83 to 4.27) but found an even stronger association for osteoarticular conditions (OR: 3.37 to 5.10) (10).

1.4. CO- AND MULTIMORBIDITY MEASUREMENT TOOLS

1.4.1. THE CHARLSON COMORBIDITY INDEX

There are different instruments to assess multimorbidity and the choice of instrument depends on the type of data available, study population and specific outcome of interest (23). The Charlson Comorbidity Index (CCI) is one of the most widely used tool for comorbidity risk adjustment (24). It contains 17 disease categories, each with an associated weight based on the associated risk of mortality (Table 1). Since its original publication in 1987, it has been validated for its ability to predict mortality in various disease groups, such as cancer, renal disease, stroke, intensive care and liver disease (25). Moreover, it has been adapted and validated to assess system-centered outcomes such as healthcare utilization and healthcare costs (26). Data are either obtained through chart review, International Classification of Diseases (ICD) diagnosis codes or questionnaires (27,28). Recently, Quan et al. published a revision of the CCI, with adapted weights according to current disease prognoses (Table 1.1) (25).

Table 1.1. Conditions included in the Charlson Comorbidity Index (24), their associated weights and the adapted weights as published by Quan et al. (25)

Assigned weights for diseases	Adapted weights proposed by Quan et al.	Conditions
1	0	Myocardial infarct
1	2	Congestive heart failure
1	0	Peripheral vascular disease
1	0	Cerebrovascular disease
1	2	Dementia
1	1	Chronic pulmonary disease
1	1	Connective tissue disease
1	0	Ulcer disease
1	2	Mild liver disease
1	0	Diabetes
2	2	Hemiplegia
2	1	Moderate or severe renal disease
2	1	Diabetes with end organ damage
2	2	Any tumor
2	2	Leukemia
2	2	Lymphoma
3	4	Moderate or severe liver disease
6	6	Metastatic solid tumor
6	4	AIDS

1.4.2. ELIXHAUSER´S COMORBIDITY MEASURE

The Elixhauser´s Comorbidity Measure (ECM) is a recent and frequently applied comorbidity index (29). The Elixhauser´s comorbidity tool defines 30 comorbidities, the 17 included in the CCI and 13 new ones (Table 1.2). Administrative data are used to identify the conditions and these are treated separately or as a count (30). The ECM seems to be a better risk adjustment tool than the CCI, especially for mortality beyond 30 days (23,31–33). However, difficulty in terms of feasibility of collecting 30 comorbidities might lead investigators to use the latter rather than the ECM.

Table 1.2. Comorbidities included in the Elixhauser's Comorbidity Measure

Comorbidity	
Congestive heart failure	Lymphoma
Valvular disease	Metastatic cancer
Pulmonary circulation disorders	Solid tumour without metastasis
Peripheral vascular disorders	Rheumatoid arthritis/collagen vascular
Hypertension	diseases
Paralysis	Coagulopathy
Other neurological disorders	Obesity
COPD	Weight loss
Diabetes uncomplicated	Fluid and electrolyte disorders
Diabetes complicated	Blood loss anaemia
Hypothyroidism	Deficiency anaemia
Renal Failure	Alcohol abuse
Liver disease	Drug abuse
Peptic ulcer excluding bleeding	Psychosis
AIDS	Depression

COPD, Chronic Obstructive Pulmonary Disease

1.4.3. THE CUMULATIVE ILLNESS RATING SCALE

The Cumulative Illness Rating Scale (CIRS) (34) is a comorbidity index that was adapted to be used in geriatric populations by Miller et al. (CIRS geriatrics: CIRS-G) (35). This scale rates 14 body systems (Table 1.3) according to a five point severity scale: 0, no problem; 1, mild current problem or past significant problem; 2, moderate disability or morbidity; requires "first line" therapy; 3, severe or constant significant disability; uncontrollable chronic problem; 4, extremely severe (life threatening), end organ failure, severe impairment in function. The total score is taken into account as well as the mean severity score for the evaluated body systems. The CIRS and CIRS-G seem to be a useful comorbidity measures in clinical research due to their structure according to clinically relevant body systems (36).

Table 1.3. Body systems included in the CIRS-G

Body systems
Heart
Vascular
Hematopoietic
Respiratory
EENT (eyes, ears, nose, throat, larynx)
Upper gastrointestinal tract
Lower gastrointestinal tract
Liver
Renal
Genito-urinary (ureters, bladder, urethra, prostate, genitals)
Musculoskeletal / integument
Neurological
Psychiatric illness
Endocrine / metabolic

1.4.4.THE SELF-ADMINISTERED COMORBIDITY QUESTIONNAIRE

Comorbidity data from medical records or administrative data can be limited, due to poor quality of the documentation, lack of recent documentation and under-reporting of pre-admission conditions not relevant for the admission diagnosis (37). Therefore, obtaining information about comorbidity through self-report has gained interest in the past years. Studies have shown that patients are generally capable of giving accurate information about their current and past medical conditions (37–39). The Self-Administered Comorbidity Questionnaire (SCQ) was first published by Sangha et al. (9). It contains 12 medical conditions, and participants are asked whether they have the condition, and if so, whether they receive treatment for it and if it limits their daily activities (Table 1.4). Patients can add three additional conditions in an open-ended way. For every positive answer, one point is assigned and, as in other comorbidity measures, a total sumscore is obtained. The SCQ showed a moderate correlation with the CCI (0.55) and it is more related to QoL than the CCI (40).

Table 1.4. The Self-Administered Comorbidity Questionnaire (9)

	Do you have the problem?		Do you receive treatment for it?		Does it limit your activities?	
Problem	No (0)	Yes (1) →	No (0)	Yes (1)	No (0)	Yes (1)
Heart disease	N	Y	N	Y	N	Y
High blood pressure	N	Y	N	Y	N	Y
Lung disease	N	Y	N	Y	N	Y
Diabetes	N	Y	N	Y	N	Y
Ulcer or stomach disease	N	Y	N	Y	N	Y
Kidney disease	N	Y	N	Y	N	Y
Liver disease	N	Y	N	Y	N	Y
Anemia or other blood disease	N	Y	N	Y	N	Y
Cancer	N	Y	N	Y	N	Y
Depression	N	Y	N	Y	N	Y
Osteoarthritis, degenerative arthritis	N	Y	N	Y	N	Y
Back pain	N	Y	N	Y	N	Y
Rheumatoid arthritis	N	Y	N	Y	N	Y
Other medical problems (please write in)						
	N	Y	N	Y	N	Y
	N	Y	N	Y	N	Y
	N	Y	N	Y	N	Y

1.5. THE DISEASE BURDEN MORBIDITY ASSESSMENT

The SCQ introduced a new aspect to comorbidity measurement: the impact of comorbidity on daily life. Bayliss et al. took a step further: they created a self-report comorbidity assessment instrument that not only assesses whether comorbidities limit daily activities, but also to what extent they do so (41). They created a list of 23 common chronic conditions (Table 1.5), asking participants whether they have each condition, and for each present condition, to what extent it interferes with daily life activities, on a scale from 1 (not at all) to 5 (a lot). Scores assigned to single conditions are summed in order to get the total score, and this measure of disease severity was conceptualized as disease burden. There is no clear definition of disease burden, also called burden of ill health (42) in the literature. Usually it is defined as 'the impact of disease events on various dimensions of human life, including health' (43).

Bayliss et al. published four articles about the instrument. In the first one, an initial validation was performed in persons aged 65 years an older (41). Sensitivity and specificity relative to chart review were calculated, as well as correlations with general health status, physical functioning, depression and self-efficacy. Mean sensitivity and specificity relative to chart review was 75% and 92%, respectively. The scale showed moderate to high correlations with overall health status, physical functioning and self-efficacy and a low correlation with depression. These correlations were higher for the DBMA than for the CCI or the RxRisk score.

In a second paper, Bayliss et al. studied the relationship between barriers to self-management, including disease burden, and perceived health and physical functioning using multivariate regression modeling (44). This analysis was done in comorbid persons aged 65 years and older, with at a minimum diabetes, depression and osteoarthritis.

Disease burden was significantly associated with both perceived health and physical functioning.

In the third article that Bayliss and colleagues published about the scale, a multivariate regression was performed with disease burden as the dependent variable (45). Biopsychosocial factors and demographic variables were included as independent variables, as well as two data-based comorbidity indices (the Quan comorbidity index and the chronic disease score, CDS). In this study, the original 23-items list was reduced to 21 chronic conditions (Table 1.5). The same sample was used as in the second article (44): persons aged 65 years and older with at least three chronic conditions. Age, 'compound effects of conditions' (treatments and symptoms interfering with each other), self-efficacy, financial constraints, and physical functioning were found to be significantly associated with disease burden. These associations were not affected by the inclusion of the other comorbidity measures in the models.

In the fourth article (46), self-reported disease burden, as a subjective measure, was compared to morbidity measured using administrative data (CCI) as an objective measurement, in the way that both related to patient-reported and utilization outcomes. As in their second and third paper(44,45), a sample of adults aged 65 years or more with at least three comorbidities was used, but this time the three medical conditions could be any of a list of 10 common chronic conditions. Self-reported disease burden was more strongly associated with patient-reported outcomes, whereas morbidity measured by diagnosis codes showed a stronger association with higher utilization.

The disease burden measure was denominated the Disease Burden Morbidity Assesment (DBMA) by Poitras et al. (47). They published a paper about the validation of the French version of the DBMA (DBMA-Fv) in a primary care setting in patients aged 18 years and older in Quebec, Canada. Bayliss´ reduced 21-item list of chronic conditions was used (45), and depression was added as a 22nd item(Table 1.5). Sensitivity and specificity relative to chart review were assessed as it had been by Bayliss et al. (41), as well as the correlation with the CIRS. Test-retest reliability was assessed by repeating the questionnaire in the form of a mail survey two weeks after the first questionnaire. Sensitivity and specificity of 74% and 92% were found, very similar to those reported by Bayliss and collegues (41). Correlations with the CIRS were 0.46 and 0.56 for the first and second surveys, respectively. An intraclass correlation coefficient (ICC) of 0.86 was found, indicating high test-retest reliability.

The same research team published another paper about the DBMA-Fv. In this study, the relation between the DBMA-Fv and literacy was studied, again in a primary care population aged 18 years and older (48). In this research, a simplified version of the DBMA was used, with only 11 chronic conditions (Table 1.5). Literacy and the DBMA were associated in bivariate analyses, but no longer in multivariate analyses when controlling for age and family income.

A third study was performed using the DBMA-Fv, also in Quebec, Canada. In this research the relation between the DBMA and obstructive sleep apnea (OSA) was assessed (49). Patients were recruited from the records of a sleep laboratory and were between 30 and 75 years old. No relation between OSA in general and the DBMA was found in bivariate analysis, but for severe OSA, a positive association was found with disease burden measured with the DBMA (OR: 7.33).

Table 1.5: Chronic conditions included in the DBMA

Chronic Condition	Bayliss et al. (2005) (41)	Bayliss et al. (2009) (45)	Poitras et al. (2012) (47)	Hudon et al. (2012) (48)
Angina/coronary artery disease	x	x	x	x
Asthma	x	x	x	x
Back pain	x	x	x	x
Bronchitis, chronic/COPD	x	x	x	x
Cancer (within the past 5 yrs)	x	x	x	
Cholesterol, elevated	x	x	x	x
Colon problem (e.g., diverticulitis, irritable bowel)	x	x	x	
Congestive heart failure	x	x	x	x
Depression			x	
Diabetes	x	x	x	x
Hard of hearing	x	x	x	
Hypertension	x	x	x	x
Kidney disease	x			
Nerve condition	x			
Osteoarthritis	x	x	x	x
Osteoporosis	x	x	x	
Overweight	x	x	x	x
Poor circulation (e.g., peripheral vascular disease)	x	x	x	
Rheumatic disease, other	x	x	x	x
Rheumatoid arthritis	x	x	x	
Stomach problem (e.g., gastritis, peptic disease)	x	x	x	
Stroke	x	x	x	
Thyroid disorder	x	x	x	
Vision problem	x	x	x	

COPD, Chronic Obstructive Pulmonary Disease

1.6. SCALE VALIDATION: CLASSICAL TEST THEORY AND RASCH ANALYSIS

1.6.1. SCALE VALIDATION

As explained in the previous chapter, initial validations of the DBMA were performed by Bayliss et al. (41,44–46) and Poitras et al (47). Criterion validity relative to chart review was assessed (41,47), as well as convergent validity with other self-reported outcomes as physical functioning, perceived health and depression (41), test-retest reliability(47) and concurrent validity with other comorbidity measures (41,45,47).

A next step, essential to guarantee the quality of a scale, is to validate it following a very precise psychometric or clinimetric methodology (50). This is done by means of the statistical analysis of different psychometric attributes, following a set of norms and standards based on scientific methods and theories of health measurement. In this sense, there are two main approaches: the classical test theory (CTT) and the item response theory (IRT), which includes Rasch analysis (51).

1.6.2. CLASSICAL TEST THEORY

CTT is a quantitative approach to test the validity and reliability of a scale. It can be traced back to Spearman at the beginning of the 20th century, who introduced the separation of an observed score into a true score and an error, and the estimation of the reliability of observed scores (52). CTT assumes that each observed score on a scale is a combination of a true underlying score and an unsystematic or random error (51). According to this theory, every person has a true score that would be obtained if there were no measurement errors. However, the true score is never shown by the scale, only an

observed score, which is assumed to be the true score plus some error. The following properties are studied in the CTT approach (50,53):

- Feasibility: The applicability of the instrument in the intended context
- Acceptability: The extent to which the instrument is acceptable in the target population.
- Scaling assumptions: The extent to which single items are related to the total score
- Reliability: The degree to which the instrument is free of random error:
    - Internal consistency: the extent to which the items measure the same construct
    - Reproducibility: Stability of the scores in time or among different evaluators
- Validity: The degree to which the instrument measures the constructs it is supposed to measure:
    - Construct validity: The degree in which the scores of the instrument are consistent with hypotheses about the construct
        - Convergent/ Divergent validity
        - Known-groups validity
        - Predictive validity
    - Criterion validity: Relation of the scale with a gold standard
        - Concurrent validity: Relation of the scale with an existing test
- Dimensionality: the existence or not of subscales (dimensions) within the scale

1.6.3. RASCH ANALYSIS

IRT can be traced back to 1927 when the Law of Comparative Judgment was published by Louis Thurstone (54). IRT can be described as a measurement model that tries to explain the connection between observed item responses on a scale and an underlying construct (51). Stochastic models are used to generate statistical estimations of parameters that represent the locations of persons and items on a latent continuum (52).

Rasch analysis was born out of the work of Georg Rasch, a Danish mathematician, in the 1960s (55). The Rasch model assumes that there is a logistic function of the difference between the item difficulty and the person's ability , or in other words, the level of the construct being measured and the person's level of the construct (56). It is based on two basic assumptions: local independence and unidimensionality. Rasch analysis is considered by some authors as a standard for developing new instruments and assessing the quality of existing ones (57). It provides a linear measure, which, given an appropriate distribution, permits the use of parametric statistics (58).

The key difference between Rasch analysis and CTT is that the latter describes a set of data, whereas Rasch analysis aims to obtain data that fit the Rasch model (52). Rasch analysis includes the following evaluations (59):

- Fit to the Rasch model: The extent to which the data coincide with theoretical item performance according to the Rasch model
- Reliability: The degree to which the instrument is free of random error
- Undimensionality: Only one construct is measured
- Response dependency: Are items are linked in such way that the response on one item will determine the response on another?

- Category structure: The extent to which the responses to the items are consistent with the metric estimate of the underlying construct
- Scale targeting: An indication of how well targeted the items are for people in the sample
- Differential item functioning (DIF): Do different groups within the sample, despite equal levels of the characteristic being measured, respond in a different manner to an individual item?

2. OBJECTIVES

The objective of this thesis was to perform a validation study of the DBMA. To do so, three studies were performed. The first study consists of a validation according to the CTT. Furthermore, since construct validity gives important information about potential contexts a scale can be applied in, this was assessed more thoroughly in a second study. In the third study, a Rasch analysis was performed. This responds to the following objectives:

OBJECTIVE 1:

To perform an analysis of the psychometric properties of the DBMA according to the assumptions of the Classical Test Theory:

- Feasibility
- Acceptability
- Scaling assumptions
- Reliability
- Construct validity
- Exploratory factor analysis

OBJECTIVE 2:

To asses construct validity by studying:

- Known-groups validity for sex and age groups
- Convergent validity: the relation between the DBMA, patient-centered outcomes, and healthcare utilization
- Predictive validity: the association with mortality.

OBJECTIVE 3:

To perform an analysis following the Item Response Theory of the DBMA through Rasch analysis:

- Test of fit to the Rasch model
- Reliability
- Unidimensionality
- Response dependency
- Category structure
- DIF
- Scale targeting
- CTT analysis of the linear measure

The results of these three studies will be published shortly. Study 1 had already been published at the moment of this writing (60); the article is added as an appendix to this thesis . Study 2 is in process of revision for resubmission in a different journal and Study 3 was accepted for publication (61) and is also presented as an appendix.

3. METHODS

3.1. STUDY DESIGN AND SAMPLE

Data came from the Ageing in Spain Longitudinal Study, Pilot Survey (Estudio Longitudinal Envejecer en España, pilot study, ELES-PS), which included 1747 community-dwelling adults aged 50 or more living in Spain (62). In this survey, a representative sample was selected on a national geographical basis. For sampling, stratified clusters of census sections were randomly selected by autonomous region and municipality, proportionally to their population of 50 years and older. Households with a telephone line were selected at random from a commercial household telephone directory. Per household, individuals aged 50 or more were randomly selected, with post-stratification by sex and age group (50-59, 60-69, 70-79 and 80-89 years). Field work was conducted in 2011.

The data in the ELES-PS study were collected in four stages: a telephone questionnaire (n=1747), a visit by a trained nurse (n=1531), a computer-assisted personal interviewing (CAPI) questionnaire (n=1400), and a self-administered questionnaire (n=1145). DBMA data were collected through the CAPI questionnaire, and its 1400 participants formed the sample that was used for the current work.

The DBMA was developed to be used in older adults (41). Therefore, Study 1 was performed with a subsample of the persons aged 65 years or older that answered the CAPI questionnaire (n=707). Because we used the percentage of missing values in the CTT approach, persons that did not answer the DBMA (completely) were included. These 82 persons that had missing values for the DBMA were excluded in Study 2 since they would not contribute to the analyses, resulting in a sample of 625 persons. Study 3 was performed with the whole study sample (n=1400). However, since analysis with samples larger than 300 could result in statistically significant deviations from the Rasch model of

otherwise well-fitting items, a random subsample of 300 was taken for the Rasch analysis (63–66).

The 1747 participants in the ELES-PS form a representative sample of the Spanish population. However, people from the Basque region were overrepresented in the sample. Due to the complex design and this overrepresentation, analyses carried out with these data base must take into account the design variables as well as the weighing factors, if possible. Since neither CTT nor Rasch analysis can be performed correcting for complex samples or weights, this was only done in the second study.

The data on mortality from all causes were obtained after a follow-up time of four years, and were extracted from the Spanish National Death Index (*Índice Nacional de Defunciones*), which includes all deaths registered in Spain since 1987. Persons were searched automatically and manually by full name, sex and date of birth. The register does not provide information about the cause of death, only the date of death is provided.

The ELES-PS study was approved by the Ethics Committee of the Spanish National Research Council. Informed consent was obtained from all individual participants included in the study. Since obtaining mortality data from the Spanish National Death Index was not among the initial objectives of the ELES study, a second approval was obtained from the Ethics Committee of the Institute of Health Carlos III for Study 2.

3.2. ASSESSMENTS

3.2.1. THE DBMA

The DBMA, first described by Bayliss et al. (41), consists of a self-report questionnaire in which participants rate the disease burden caused by a number of medical conditions, if present. Patients are asked to what extent conditions interfere with daily activities, on a five-point scale from 1 (not at all) to 5 (a lot). Conditions not present are scored zero. As in other studies (Table 1.5), the original DBMA's 23-item list was adapted by selecting 21 common chronic conditions based on the conditions used in other multimorbidity indices (9,41,67–69). From the multimorbidity indices for which in the validation studies analysis per chronic condition had been performed, those conditions that specifically predicted mortality, hospitalization or future handicaps and those that showed a transversal association with physical functioning were selected (67–69). In case analysis per condition had not been performed, all conditions included in these multimorbidity indices were selected (9,41). As a criterion, only conditions selected from more than one index were included. A few exceptions were made: liver diseases were not included because of their low prevalence in older adults; and urinary tract conditions, anxiety and memory related disorders were added because of their high prevalence in this population. The DBMA, as used in the ELES-PS study, is shown in Table 3.1.

The DBMA is a measure of disease burden but can also be applied as a disease count, since it asks for the presence of 21 chronic conditions. Other authors applied the scale for this purpose (70) and so did we in one of the analyses in Study 1. However, in the present work, when using the term 'DBMA score' it always refers to disease burden, unless stated otherwise.

Table 3.1. The DBMA as used in the ELES-PS study

			In case of an affirmative answer, ask:
		Do you have the condition?	On a scale of 1 to 5, can you tell me whether the condition has limited you in your usual activities?
		Yes/No	1 (not at all) 2 3 4 5 (a lot)
1	Hypertension	1 2	1 2 3 4 5
2	Myocardial infarction	1 2	1 2 3 4 5
3	Heart failure	1 2	1 2 3 4 5
4	Angina	1 2	1 2 3 4 5
5	Circulation problems/ intermittent claudication	1 2	1 2 3 4 5
6	Osteoarthritis	1 2	1 2 3 4 5
7	Rheumatoid arthritis	1 2	1 2 3 4 5
8	Asthma	1 2	1 2 3 4 5
9	COPD/emphysema	1 2	1 2 3 4 5
10	Diabetes	1 2	1 2 3 4 5
11	Gastric/duodenal ulcer	1 2	1 2 3 4 5
12	Kidney disease	1 2	1 2 3 4 5
13	Depression	1 2	1 2 3 4 5
14	Anxiety	1 2	1 2 3 4 5
15	Cerebral embolism/stroke	1 2	1 2 3 4 5
16	Cancer	1 2	1 2 3 4 5
17	Osteoporosis	1 2	1 2 3 4 5
18	Memory disorders	1 2	1 2 3 4 5
19	Parkinson's disease	1 2	1 2 3 4 5
20	Chronic back pain	1 2	1 2 3 4 5
21	Urinary tract problems (prostate, bladder)	1 2	1 2 3 4 5

COPD, Chronic Obstructive Pulmonary Disease

3.2.2. OTHER ASSESSMENTS

To screen for depression, the self-administered questionnaire of the ELES-PS included the dichotomous 10-item Center for Epidemiologic Studies Depression Scale (CES-D) (71). It contains 10 questions with 'yes/no' response categories, asking about the feelings of the respondent in the past week (Table 3.2). A score of 1 is assigned to every positive answer for depression. A sum score of 3 was used as a cut-off point for depression (72). Previous studies found support for this short version of the CES-D to be as reliable as the original CES-D, with a Cronbach´s alpha of 0.80, and to show satisfactory convergent validity with the Composite International Diagnostic Interview (sensitivity and specificity of 84% and 64%, respectively) (71,72).

Table 3.2. The Center for Epidemiological Studies Depression Scale, 10-item version (71)

	yes	no
I felt depressed	1	0
I felt that everything I did was an effort	1	0
My sleep was restless	1	0
I was happy	0	1
I felt lonely	1	0
People were unfriendly	1	0
I enjoyed life	0	1
I felt sad	1	0
I felt that people disliked me	1	0
I could not get going	1	0

QoL was assessed through the Personal Wellbeing Index (PWI) (73), which provides a measure of general QoL or wellbeing. The PWI was included in the CAPI-questionnaire. Respondents are asked to grade, on a scale of 1 to 10, their satisfaction with 7 life dimensions: standard of living, personal health, achieving in life, personal relationships, personal safety, community-connectedness and future security. Total subscores were lineally transformed into a 0-100 scale (58), and higher total scores indicate better QoL. Previous research found support for the validity and reliability of this linear measure in older adults, correlating moderately with 'satisfaction with life' and showing a person separation index (PSI) of 0.91 (58).

The CAPI questionnaire included a question about perceived health, asking the participants to grade their satisfaction with their health status on a scale from 1 (very bad) to 5 (very good). This variable was used in Study 2, and dichotomized into very good/ good vs. acceptable/poor/very poor (74). However, the PWI also includes a dimension about personal health answered on a 0-10 rating scale. Because of this broader response scale, this dimension was used as a measure of perceived health in Studies 1 and 3.

A 24-item list of different basic and instrumental activities of daily living, as used in the Health and Retirement Study (75), was included in the CAPI questionnaire as a measure of physical functioning (Table 3.3). Included activities are getting dressed, walking 100 meters and making phone calls, among others. Participants were asked whether they experienced difficulties in performing them, on a scale from 1 (always) to 4 (never). Scores were summed to get a measure of physical functioning. In Study 2, this variable was dichotomized into no disability (score of 96, which was the maximum score) vs. any level of disability (scores <96)(10).

Table 3.3. The 24-item list of different basic and instrumental activities of daily living as included in the ELES-PS CAPI questionnaire

Do you have any difficulty with…		Always	Sometimes	Almost never	Never
1	Walking 100m	1	2	3	4
2	Walking 1000m	1	2	3	4
3	Sitting for about two hours	1	2	3	4
4	Getting up from a chair after sitting for long periods	1	2	3	4
5	Climbing several flights of stairs without resting	1	2	3	4
6	Climbing one flight of stairs without resting	1	2	3	4
7	Stooping, kneeling, or crouching	1	2	3	4
8	Reaching or extending your arms above shoulder level	1	2	3	4
9	Pulling or pushing large objects like a living room chair	1	2	3	4
10	Lifting or carrying weights over 10kg, like a heavy bag of groceries	1	2	3	4
11	Picking up a 5 cents coin from a table	1	2	3	4
12	Dressing, including putting on shoes and socks	1	2	3	4
13	Walking across a room	1	2	3	4
14	Bathing or showering	1	2	3	4
15	Eating, such as cutting up your food	1	2	3	4
16	Getting in or out of bed	1	2	3	4
17	Using the toilet, including getting up and down	1	2	3	4
18	Using a map to figure out how to get around in a strange place	1	2	3	4
19	Preparing a hot meal	1	2	3	4
20	Shopping for groceries	1	2	3	4
21	Making phone calls	1	2	3	4
22	Taking medications	1	2	3	4
23	Work around the house or yard	1	2	3	4
24	Managing your money, such as paying your bills and keeping track of expenses	1	2	3	4

The CAPI questionnaire contained the Scale of Positive and Negative Experience (SPANE) to assess affect balance. (Table 3.4) (76). This 12-item questionnaire includes six items that assess positive feelings and six items for negative feelings. Respondents are asked to report how much they experienced each feeling in the past month, on a scale from 1 (very rarely or never) to 5 (very often or always). Scores for negative feelings are subtracted from the positive feeling total score, resulting in a total scale from -24 (unhappiest) to 24 (happiest) to obtain a measure of affect balance. Diener et al (76) found support for the validity and reliability of the scale, converging well with other measures of emotions and affective well-being, and showing a Cronbach´s alpha of 0.88.

Table 3.4. The Scale of Positive and Negative Experience as used in the ELES-PS (76)

	Very Rarely or Never	Rarely	Sometimes	Often	Very Often or Always
Positive	1	2	3	4	5
Negative	1	2	3	4	5
Good	1	2	3	4	5
Bad	1	2	3	4	5
Pleasant	1	2	3	4	5
Unpleasant	1	2	3	4	5
Happy	1	2	3	4	5
Sad	1	2	3	4	5
Afraid	1	2	3	4	5
Joyful	1	2	3	4	5
Angry	1	2	3	4	5
Contented	1	2	3	4	5

Two measures were included in the CAPI questionnaire to assess the use of healthcare resources. In the first, the numbers of visits in the past month to the primary care center (general practitioner, nurse), physiotherapist or medical specialist were summed, to get a measure of primary and outpatient care utilization. To assess the use of hospital care resources, the visits to the emergency department, 'day hospital' and hospital admissions in the past year were summed. Both variables were dichotomized, into use vs. no use of healthcare resources in the corresponding period of time.

3.3. STATISTICAL ANALYSES

Statistical analyses were performed in Stata 12 for Windows, unless stated otherwise.

3.3.1. STATISTICAL ANALYSIS IN STUDY 1: CTT

The following psychometric properties were examined: feasibility, acceptability, scaling assumptions, reliability and construct validity. Feasibility was assessed by determining the percentage of missing values per item and the percentage of computable scores for the total scale, considering acceptable scores <10% and >90%, respectively (77). Acceptability was explored by comparing possible and observed scores and assessing mean-to-median difference for the total scale (criterion, <10% of the scale range) as well as floor and ceiling effects (<15%) and skewness (-1 to 1) (78).

Scaling assumptions were determined through the item-total corrected correlation (ITCC) for each item (criterion r≥0.40) (79). Reliability was assessed through internal consistency (Cronbach's alpha) and the item homogeneity index (criteria: $\alpha \geq 0.70$ and $r \geq 0.30$, respectively) (77).

For convergent validity, we expected self-reported disease burden to be negatively associated with perceived health, physical functioning and QoL and to find a positive association with depression (10,80,81). This was calculated through Spearman's rank correlation coefficients due to the non-normal distribution of the DBMA. Correlation coefficients were interpreted following Cohen's conventions, considering correlations ≥0.5 as large, 0.5-0.3 as moderate and 0.3-0.1 as having a small magnitude (82).

Spearman's rank correlations were repeated using the self-reported number of conditions instead of disease burden, to evaluate the added value of assessing the impact of conditions on daily life. These correlations were compared using a Fisher r-to-z transformation (83). Because the DBMA had a higher number of missing values than the disease count variable, those cases with missing values for DBMA were excluded in this analysis to make the two variables comparable.

Known-groups validity was examined comparing disease burden by sex and age groups (Wilcoxon rank-sum test). Since multimorbidity is more frequent in both women and older people, we hypothesized to find significantly higher scores in these groups than in men and younger participants (84).

Dimensionality and factor structure were explored through exploratory factor analysis, using a principal axis factoring method with oblimin rotation. The number of extracted factors was determined according to eigenvalues (>1) and visual inspection of the screeplot, taking the 'elbow' as the point of separation (85).

3.3.2. STATISTICAL ANALYSIS IN STUDY 2: CONSTRUCT VALIDITY

Due to overrepresentation of the Basque Country in the sample, analyses were weighted to the underlying population distribution and accounted for the effect of stratification and clustering. Therefore, descriptive statistics are presented with are presented with standard errors (SE) instead of standard deviations (SD). Not all statistics allowed the correction for the complex design or weights; for those cases where correction was not possible, this is stated in the text.

In Study 1, known-groups validity was assessed by sex and age group (<75 years vs. ≥75 years). In Study 2, this was studied for disease prevalence and disease burden scores separately. First, differences in the mean number of present conditions per person were studied by sex and age group. Since the Wilcoxon rank-sum test does not allow weights or correction for complex design, p-values for the differences in the mean number of present conditions were obtained through a Somers' D analysis (86) weighted for the population distribution. Furthermore, an analysis of the differences in disease prevalence and disease burden scores was performed for single conditions. Significance of the differences in disease prevalence by groups was assessed with Chi-square tests, whereas significance of the difference in disease burden scores was tested with Somers' D tests.

Study 1 assessed convergent validity with perceived health, physical functioning, QoL and depression. These variables were included in the convergent validity assessment in study 2 as well, except for depression, which was excluded because of a very high proportion of missing values (24.5%). Affect balance and the use of healthcare utilization were added since we expected them to be related to the DBMA as well (46,87).

Bivariate linear regression models were used to determine which of the following independent variables should be included in the multivariate linear regression model: age,sex, patient-centered variables (perceived health, physical functioning, QoL, affect balance) and system-centered outcomes (primary/outpatient care and hospital care use). Likelihood ratio statistics were applied beforehand to decide whether variables should be used as continuous, categorical or dichotomic variables: age, QoL and affect balance were used as continous variables and perceived health, physical functioning and the use of outpatient and hospital care were dichotomized.

Independent variables that were significant at a $p \leq 0.15$ level in the bivariate analysis were considered for inclusion in the multivariate model (45). Because of the skewed distribution of the DBMA in our sample, a generalized linear model with gamma distribution and log link was used for the bivariate and multivariate regression models.

For variables that were significant in the bivariate analysis and no longer in the multivariate regression, confounding factors were assessed to investigate which variables caused this change in significance. This was done by adding potential confounding variables individually to the bivariate models of the non-significant variables. A change-in-estimate (CIE) criterion of 20% was used (88). In order to be considered as a confounder, variables should be associated with both the independent and the dependent variable. Associations with the independent variable were studied with linear regression models (regression coefficient).

This study used the Cox proportional hazards model to examine the performance of the DBMA as a predictor of survival, after adjusting for age and sex. The proportional hazards assumption was confirmed graphically. Besides, the comparison of survival between three DBMA categories (low, medium and high) was done using Kaplan-Meier curves, taking as

low DBMA the scores within the first quartile in the used sample (≤2) and as high the scores within the last quartile (≥11). This analysis does not allow correction for complex samples, so only the weights were taken into account.

A logistic regression model, with mortality as a dependent variable and DBMA as an independent variable, adjusted for age and sex, was used to calculate a Receiver Operating Characteristic (ROC) curve. The area under the ROC curve (AUC) was used to assess how well the DBMA predicted mortality, with the following cut-off points: non-predictive (AUC =0.5), less predictive (0.5<AUC<0.7), moderately predictive (0.7≤AUC<0.9), highly predictive (0.9≤AUC<1) and perfect prediction (AUC =1) (89).

3.3.3. STATISTICAL ANALYSIS IN STUDY 3: RASCH ANALYSIS

The Rasch model is a mathematical expression that assumes that a response is a logistic function of the difference between the item difficulty and the respondent's ability (55). In our case, item difficulty refers to the level of disease burden measured by the item on the constructs continuum, whereas the respondent's ability reflects how much burden a person experiences.

Rasch analysis was performed using RUMM 2030 (90). Differences between thresholds were not expected to be equal across items, so the Masters' Partial Credit polytomous model was chosen (91), which was confirmed by a significant likelihood ratio statistic. Test of fit to the Rasch model, reliability, unidimensionality, response dependency, category structure, DIF and scale targeting were studied (59). These parameters were assessed in an iterative way, making model modifications until an adequate fit was achieved.

Fit to the Rasch model was tested by comparing the observed data with the theoretical item performance according to the Rasch model. The item-trait interaction statistic, reported as a chi-square, needs to be non-significant (59). Item and person summary fit statistics should follow a normal distribution with a mean and SD of 0 and 1, respectively. Individual item and person fit residuals should be within the ±2.5 range and chi-square differences for items and persons should be non-significant with Bonferroni correction for number of items (92).

Reliability was determined with the person separation index (PSI), which is interpreted similarly to Cronbach´s coefficient alpha: a minimum value of 0.70 for group comparisons is recommended and 0.90 for person comparisons (59). The PSI was also obtained in

RUMM2020 since algorithms derived from this program provide reliability results less influenced by extreme values, missing data, and floor and ceiling effects than those obtained with RUMM2030 (93).

Unidimensionality was tested through a Principal Component Analysis (PCA) of the residuals (94). This analysis defines two subsets of items, positively and negatively correlated with the first residual factor, and the differences in these estimates for each person are compared with a t-test. The differences in estimates are supposed to be normally distributed, so the percentage of these tests outside the range of -1.96 to 1.96 should not exceed 5% in a binomial test (59).

Response dependency was assessed through the residual correlation index and a correlation of > 0.30 was taken as an indication of local dependency (58). This can be illustrated with the following example (59): if you include two walking items in a scale, one asking if a person can walk a kilometer without difficulty, and one asking if a person can walk 100 meters, a person that can walk one kilometer will always answer that he or she can walk 100 meters as well. Related items should be combined in a subtest, so in the example above, one walking item could be made with response items relative to the walking distance.

Response category structure was explored through category probability curves, and in case of disordered thresholds, items were rescored by collapsing adjacent categories. A threshold is the point of equal response probably between two adjacent response categories.

DIF examines whether different groups within the sample, despite of equal levels of the characteristic being measured, respond in a different manner to an individual item (95).

DIF was studied for age (< 65 years vs. ≥65 years, 65 was the median value in our sample), sex and educational level (primary school or less vs. more than primary school). The DIF analysis was done through an analysis of variance (ANOVA) with Bonferroni correction. In case DIF was identified, this was further analyzed through a top-down purification approach (95). In this approach, items are divided into two groups, according to the presence of absence of DIF, and two testlets (or superitems) are created. If the testlet formed by the items with DIF does not present DIF itself, then it is considered to cancel out (96).

Scale targeting was assessed to analyze whether the sample and items covered all levels of the construct continuum. This was done through visual inspection of a graphic showing the distribution of persons and items along the construct.

Once fit to the Rasch model was achieved, disease burden scores of the total sample were used to calculate a linear measure, on a logit scale, which was converted into a 0-47 range using a linear transformation. In order to compare the subsample of 300 and the rest of the sample (n=1100), a paired-sample t-test was done, comparing the logit estimation of the two samples (300 vs. 1100) for each raw-score. Anchor values of the sample of 300 were used to fix item estimations of the other sample. In addition, a DIF analysis by sample (300 vs. 1100) was performed.

Psychometric attributes of the linear measure according to the CTT were analyzed in Stata: mean-to-median differences (criterion,<10%), floor- and ceiling effects (<15%) and skewness (-1 to 1) were calculated for acceptability (78). Construct validity was assessed through known-groups validity for sex and age (<65 years vs. ≥ 65 years) and convergent validity with other health outcomes. We hypothesized to find higher disease burden scores for women (97) and in the highest age group (98), which was studied with a Wilcoxon

rank-sum test due to the non-normal distribution of the linear measure. For convergent validity, Spearman's rank correlations were calculated with physical functioning, depression (CES-D total score), QoL (PWI), and perceived health (PWI item 2: personal health). Moderate to high correlations (r>0.30) were expected (82).

A relative precision analysis was performed in order to compare the ability of the Rasch-based score in distinguishing groups relatively to the raw summative-based score (99). This was done for sex and age groups (<65 years vs.≥ 65 years). Relative precision was calculated as the ratio of pairwise Z statistics (the linear measure Z-statistic divided by the raw score Z statistic) (100), and a bootstrap method was applied in order to obtain confidence intervals (CI) for relative precision statistics. For each patient group comparison, a total of 1000 bootstrap samples (with replacement) were drawn, and F statistics and relative precision values were calculated for each resampling. The 25th and the 975th estimates of these values were taken as the limits of the 95% CI (101). Rasch analysis takes into account observations with missing values when calculating the linear measure. In order to be able to calculate the relative precision, the studied sample sizes should be equal, thus observations with missing values were excluded in the latter analysis.

4. RESULTS

4.1. DESCRIPTIVE STATISTICS OF SOCIODEMOGRAPHIC DATA AND APPLIED RATING SCALES

Table 4.1 presents the characteristics of the total study sample (n=1400), the Rasch analysis subsample (n=300), and the sample used in Study 1 (n=707). Since the sample used in the Rasch analysis is supposed to be comparable with the whole sample, this column is presented second, followed by the data of Study 1.

As stated, statistical analyses in Study 2 were corrected for the complex design and weighted for the sampling method. Therefore, the results are presented in a separate table (Table 4.2). In this table, descriptive statistics of the total sample of Study 2 are presented, as well as the sample used in the multivariate regression and Spanish national data.

Table 4.1. Characteristics of the total study sample (n=1400), the Rasch analysis subsample (n=300), and the sample of 65 years and older (Study 1, n=707)

		Total sample (n=1400)		Sample for Rasch analysis (n=300)		Sample for Study 1 (n=707)	
Characteristic		n	(%)	n	(%)	n	(%)
Sex	Men	625	(44.6)	132	(44.0)	304	(43.0)
	Women	775	(55.4)	168	(56.0)	403	(57.0)
Education	Less than primary	480	(34.3)	98	(32.7)	365	(51.6)
	Primary	313	(22.4)	77	(25.7)	137	(19.4)
	Secondary	298	(21.3)	61	(20.3)	83	(11.7)
	University	309	(22.1)	64	(21.3)	122	(17.3)
Living area	< 10.000 inhabitants	315	(22.5)	70	(23.3)	133	(18.8)
	10.000-100.000 inhabitants	502	(35.9)	103	(34.3)	273	(38.6)
	100.000-500.000 inhabitants	385	(27.5)	90	(30.3)	201	(28.4)
	> 500.000 inhabitants	198	(14.0)	37	(12.3)	100	(14.1)
Marital status	Single	75	(5.4)	17	(5.7)	31	(4.4)
	Married/ living with partner	1014	(72.4)	225	(75.0)	439	(62.1)
	Widowed	244	(17.4)	44	(14.7)	211	(29.8)
	Divorced/separated	67	(4.8)	14	(4.7)	26	(3.7)
CES-D†	Depression	297	(21.2)	78	(26.0)	159	(22.5)
	No depression	797	(56.9)	160	(53.3)	366	(51.8)
	Missing	306	(21.9)	62	(20.7)	182	(25.7)
Characteristic, range		Mean	±SD	Mean	±SD	Mean	± SD
Age in years		65.5	± 10.4	64.96	± 10.3	74.2	± 6.6
PWI, 0-100		74.9	± 11.1	73.3	± 12.5	75.3	± 11.1
Physical functioning, 24-96		91.1	± 9.6	90.7	± 10.1	88.3	± 11.8
Satisfaction with health, 0-10		7.2	± 3.9	6.8	± 1.9	7.1	± 5.3
Self-reported number of conditions, 0-21		2.5	± 2.3	2.6	± 2.3	3.2	± 2.4
DBMA raw score, 0-105		5.3	± 6.4	5.4	± 6.2	6.8	± 7.1

CES-D, Center for Epidemiologic Studies Depression Scale; DBMA, Disease Burden Morbidity Assessment; PWI, Personal Wellbeing Index; SD, Standard Deviation.

†cut-off point: 3 out of 10

Table 4.2. Characteristics of the study sample in Study 2: total sample, sample included in the multivariate regression (MVR), and Spanish national data

		Total sample (n=625)		Sample MVR (n=503)		National data ‡
Characteristic		n (%)†		n (%)†		(%)
Sex	Men	280	(44.6)	238	(47.0)	(43.0)
	Women	345	(55.4)	265	(53.0)	(57.0)
Education	Less than primary	310	(49.8)	245	(48.9)	
	Primary	126	(20.2)	100	(19.9)	
	Secondary	73	(10.7)	58	(10.9)	
	University	116	(19.4)	100	(20.3)	
Living area	< 10.000 inhabitants	125	(21.8)	104	(22.0)	
	10.000-100.000 inhabitants	237	(35.4)	186	(34.7)	
	100.000-500.000 inhabitants	165	(23.3)	136	(24.7)	
	> 500.000 inhabitants	98	(19.5)	77	(18.6)	
Marital status	Single	29	(4.9)	23	(4.9)	
	Married/ living with partner	394	(60.8)	329	(63.3)	
	Widowed	179	(29.3)	135	(28.1)	
	Divorced/separated	23	(4.2)	16	(3.7)	
Perceived health	Very good/good	353	(57.4)	299	(57.1)	(44.2)
	Acceptable/poor/very poor	262	(42.6)	204	(42.9)	(55.8)
	Missing §	10				
Functional status	No disability	249	(41.8)	221	(43.4)	
	Disability	347	(58.2)	282	(56.6)	
	Missing §	29				
Primary/outpatient care past month	Yes	405	(67.3)	330	(67.6)	
	No	220	(32.7)	173	(32.4)	
Hospital care past year	Yes	171	(29.5)	139	(29.3)	
	No	454	(70.5)	364	(70.7)	
Mortality	Living	590	(94.5)	473	(94.0)	
	Deceased	35	(5.5)	30	(6.0)	(12.0) ¶
Characteristic, range		n	Mean (SE)	Mean	(SE)	Mean
Age in years		625	73.9 (0.4)	73.7	(0.4)	75.6
PWI, 0 - 100		540	75.4 (0.6)	75.5	(0.6)	
SPANE, -24 - 24		622	12.9 (0.3)	13.1	(0.4)	
Self-reported number of conditions, 0-21		625	3.22 (0.1)	3.15	(0.1)	
DBMA, 0 - 105		625	7.5 (0.4)	7.5	(0.4)	

†Unweighted counts and weighted percentages.
‡ Data obtained from the Spanish Population and Housing Census 2011 (102) and the Spanish National Health Survey 2011-2012 (103).
§ In order to make the columns comparable, missing values were not taken into account when calculating percentages
¶ Expected 4 years mortality proportion in our total sample calculated with national mortality data 2013 (104).

DBMA, Disease Burden Morbidity Assessment; MVR, Multivariate Regression; PWI, Personal Wellbeing Index; SE, Standard Error; SPANE, Scale of Positive and Negative Experience

As shown in Table 4.1, mean age in the total sample was 65.5 years and 55.4% were women. The mean number of self-reported conditions was 2.5, with a mean DBMA score of 5.3. The distribution of the DBMA in this sample is shown in Figure 4.1. DBMA scores ranged 0-55 and there was an important floor effect: 253 of participants (18.0%) had a DBMA score of 0. In the sample aged 65 years and older (Study 1), the mean age was 74.2 and it showed a slightly higher percentage of women (57.0%). Among these participants, the mean number of conditions was 3.2 and the mean DBMA score was 6.8. The distribution in this sample (Figure 4.2) ranged 0-41 and still showed a floor effect (10.6%). Many participants had a DBMA score of 4 or less.

The participants in Study 2 were 65 years and older as well. In this subsample, a mean age was 73.9 years was found after correction for complex samples (Table 4.2). 55.4% of the participants were women. Mean DBMA score was 7.5 and after four years, 35 of the participants had died (5.5%, weighted percentage). In the total sample of persons aged 65 years and older (N=707), although not displayed in the tables, a total of 43 persons had died (6.6%, weighted percentage). Of the persons included in the sample of Study 2, 57.4% declared to be in good or very good health. In the Spanish National Health Survey, a percentage of 44.2% was found. The expected percentage of deceased persons after 4 years in the sample was 12% according to the estimation made by applying mortality rates from the National Statistics Institute to the used sample.

Figure 4.1. The distribution of the DBMA in the whole study sample (n=1277)

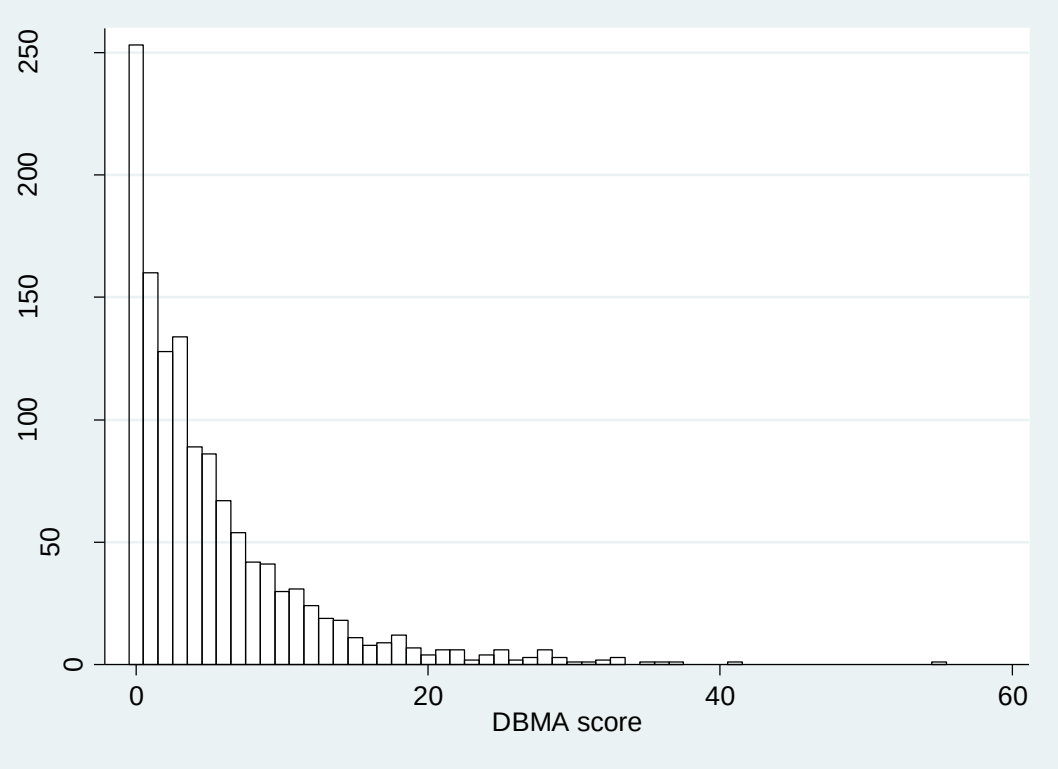
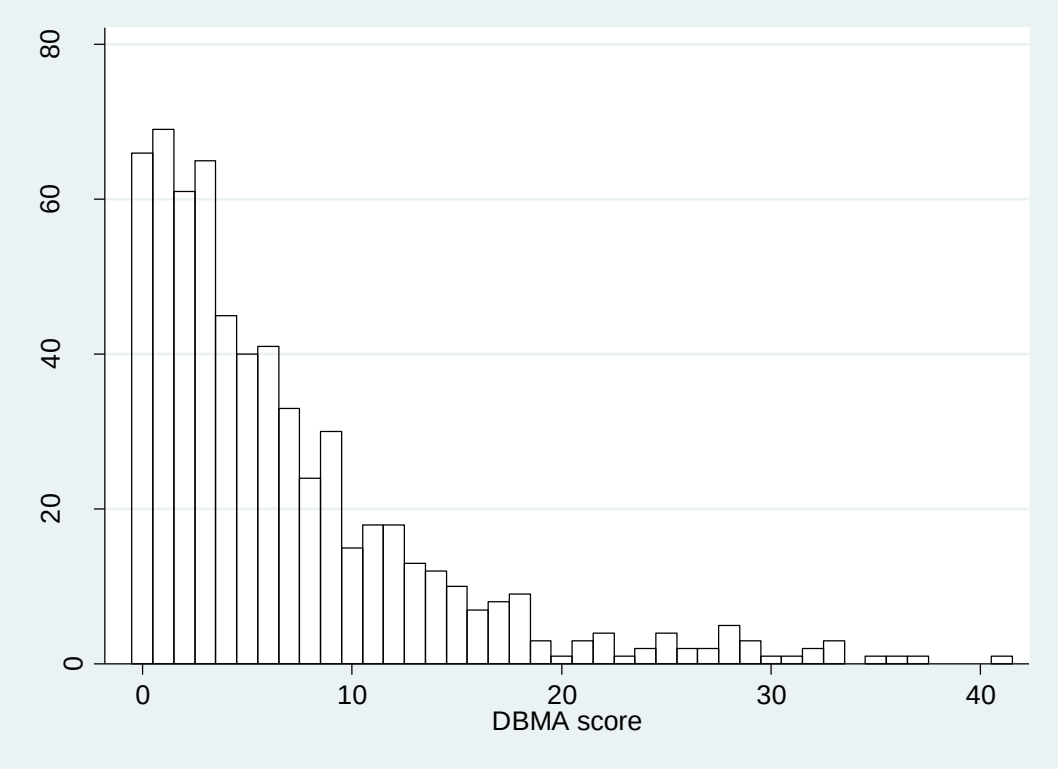


Figure 4.2. The distribution of the DBMA in the sample aged 65 years and older (n=625)



4.2. RESULTS OF STUDY 1: CTT ANALYSIS

4.2.1. FEASIBILITY, ACCEPTABILITY, SCALING ASSUMPTIONS, RELIABILITY AND CONSTRUCT VALIDITY

Tables 4.3.1 and 4.3.2 present the prevalence of conditions, disease burden per condition and scale validation data in the analysed sample. These analyses were not corrected for complex samples so the prevalences of conditions will be discussed in Study 2. All items had less than 4% missing responses and there were 88.4% computable scores. The observed and possible range was 0-5 for all items, except for Parkinson´s disease (0-4). All items had a median score of 0, The median for the total scale was 5, with a mean-median difference of 1.7%. For all items, floor effects were above 50% and ceiling effects were below 3%. When only studying the present conditions, there was still a floor effect, but less pronounced, ranging from 14.1% to 66.8% (Table 4.4). The conditions with the highest floor effect when present were hypertension (66.8%), kidney disease (64.3%) and gastric/duodenal ulcer (63.9%). Two conditions showed a ceiling effect when present: memory disorders (15.8%) and chronic obstructive pulmonary disease (COPD)/emphysema (10.9%).

Skewness was 1.80. ITCC was low for all conditions (range: 0.10-0.49), with only six conditions meeting the criterion of ≥0.40: osteoarthritis (0.43), intermittent claudication (0.43), rheumatoid arthritis (0.49), chronic back pain (0.45), depression (0.45) and anxiety (0.44). Cronbach's alpha was 0.72, and the item homogeneity index was 0.09.

Data on convergent validity are shown in Table 4.5. The DBMA had a Spearman's correlation of -0.56 with physical functioning and perceived health, -0.41 with PWI and 0.41 with depression ($p<0.001$ for all). All correlations were significantly stronger for the

DBMA than for the number of diseases. Women had higher mean DBMA scores than men (8.4 vs. 4.9, $p<0.001$). Disease burden scores increased significantly with age with a mean score of 6.1 in the <75 years of age group and 7.7 in participants aged 75 years and older ($p<0.001$).

Table 4.3. Prevalence of conditions, disease burden per condition and scale validation data in the analysed sample (n=707)

	Self-reported conditions				Self-reported disease burden		DBMA scores used for scale validation†					
Medical Condition	Missing	(%)	Condition present	(%)	Mean	(SD)	Missing	(%)	Mean	(SD)	Observed range	Floor Effect (%)	Ceiling Effect (%)	ITCC
1 Hypertension	2	(0.3)	339	(48.1)	1.5	(0.9)	19	(2.7)	0.7	(1.0)	0-5	53.2	0.3	0.24
2 Osteoarthritis	2	(0.3)	328	(46.5)	2.7	(1.2)	22	(3.1)	1.2	(1.6)	0-5	55.0	2.9	0.43
3 Circulation problems/ intermittent claudication	4	(0.6)	159	(22.6)	2.2	(1.2)	13	(1.8)	0.5	(1.1)	0-5	78.4	0.6	0.43
4 Rheumatoid arthritis	7	(1.0)	155	(22.1)	2.8	(1.2)	17	(2.4)	0.6	(1.3)	0-5	79.0	1.7	0.49
5 Chronic back pain	3	(0.4)	155	(22.0)	2.9	(1.2)	16	(2.3)	0.6	(1.3)	0-5	79.5	1.9	0.45
6 Depression	5	(0.7)	130	(18.5)	2.4	(1.3)	12	(1.7)	0.4	(1.1)	0-5	82.3	1.0	0.45
7 Urinary tract problems (prostate, bladder)	3	(0.4)	127	(18.0)	2.2	(1.2)	11	(1.6)	0.4	(1.0)	0-5	82.9	0.4	0.29
8 Osteoporosis	3	(0.4)	119	(16.9)	2.4	(1.3)	9	(1.3)	0.4	(1.0)	0-5	83.8	1.3	0.39
9 Diabetes	3	(0.4)	109	(15.5)	2.0	(1.1)	13	(1.8)	0.3	(0.8)	0-5	85.7	0.6	0.27
10 Anxiety	4	(0.6)	92	(13.1)	2.5	(1.3)	14	(2.0)	0.3	(0.9)	0-5	88.2	0.7	0.44
11 Cancer	11	(1.6)	76	(10.9)	2.0	(1.2)	12	(1.7)	0.2	(0.7)	0-5	89.2	0.1	0.13

DBMA, Disease Burden Morbidity Assessment; ITCC, Item-Total Corrected Correlation; SD, Standard Deviation

†Self-reported disease burden scores including value 0 if condition not present

Items are presented in descending order of prevalence

Table 4.3. Prevalence of conditions, disease burden per condition and scale validation data in the analysed sample (n=707) *continued*

		Self-reported conditions			Self-reported disease burden		DBMA scores used for scale validation†					
	Medical Condition	Missing	(%)	Condition present (%)	Mean	(SD)	Missing	(%)	Mean (SD)	Observed range	Floor Effect (%)	Ceiling Effect (%)	ITCC
12	Gastric/duodenal ulcer	6	(0.8)	74 (10.6)	1.7	(1.1)	8	(1.1)	0.2 (0.6)	0-5	89.7	0.3	0.17
13	Heart failure	4	(0.6)	74 (10.5)	2.4	(1.4)	14	(2.0)	0.2 (0.8)	0-5	90.8	0.9	0.15
14	Kidney disease	6	(0.8)	60 (8.6)	1.7	(1.1)	10	(1.4)	0.1 (0.5)	0-5	92.0	0.1	0.19
15	COPD/emphysema	5	(0.7)	56 (8.0)	2.5	(1.4)	6	(0.8)	0.2 (0.1)	0-5	92.2	0.9	0.19
16	Asthma	6	(0.8)	41 (5.9)	2.6	(1.3)	8	(1.1)	0.1 (0.7)	0-5	94.4	0.3	0.23
17	Angina	5	(0.7)	36 (5.1)	2.0	(1.2)	9	(1.3)	0.1 (0.5)	0-5	95.4	0.1	0.16
18	Myocardial infarction	5	(0.7)	32 (4.6)	2.9	(1.2)	9	(1.3)	0.1 (0.6)	0-5	96.0	0.3	0.10
19	Cerebral embolism/stroke	4	(0.6)	29 (4.1)	2.0	(1.4)	7	(1.0)	0.1 (0.5)	0-5	96.3	0.3	0.16
20	Memory disorders	4	(0.6)	26 (3.7)	2.6	(1.4)	11	(1.6)	0.1 (0.5)	0-5	97.3	0.4	0.17
21	Parkinson's disease	3	(0.4)	12 (1.7)	2.7	(1.3)	5	(0.7)	0.0 (0.3)	0-4	98.6	0	0.16
	Total						82	(11.6)	6.8 (7.1)	0-41	10.6	0	

COPD, Chronic Obstructive Pulmonary Disease; DBMA, Disease Burden Morbidity Assessment; ITCC, Item-Total Corrected Correlation; SD, Standard Deviation

†Self-reported disease burden scores including value 0 if condition not present

Items are presented in descending order of prevalence

Table 4.4. Missing data, floor and ceiling effects when conditions present

Condition	Condition present	DBMA data present	DBMA data missing (%)	Observed range	Floor effect (%)	Ceiling effect (%)
Hypertension	339	322	17 (5.0)	1-5	66.8	0.6
Osteoarthritis	328	308	20 (6.1)	1-5	22.4	6.5
Circulation problems/ intermittent claudication	159	150	9 (5.7)	1-5	42.0	2.7
Rheumatoid arthritis	155	145	10 (6.4)	1-5	20.7	8.3
Chronic back pain	155	142	13 (8.4)	1-5	14.1	9.2
Depression	130	123	7 (5.4)	1-5	33.3	5.7
Urinary tract problems (prostate, bladder)	127	119	8 (6.3)	1-5	41.2	2.5
Osteoporosis	119	113	6 (5.0)	1-5	31.9	8.0
Diabetes	109	99	10 (9.2)	1-5	43.4	4.0
Anxiety	92	82	10 (10.9)	1-5	29.3	6.1
Cancer	76	75	1 (1.3)	1-5	52.0	1.3
Gastric/duodenal ulcer	74	72	2 (2.7)	1-5	63.9	2.8
Heart failure	74	64	10 (13.5)	1-5	37.5	9.4
Kidney disease	60	56	4 (6.7)	1-5	64.3	1.8
COPD/emphysema	56	55	1 (1.8)	1-5	41.8	10.9
Asthma	41	39	2 (4.9)	1-5	28.2	5.1
Angina	36	32	4 (11.1)	1-5	50.0	3.1
Myocardial infarction	32	28	4 (12.5)	1-5	14.3	7.1
Cerebral embolism/stroke	29	26	3 (10.3)	1-5	57.7	7.7
Memory disorders	26	19	7 (26.9)	1-5	21.0	15.8
Parkinson's disease	12	10	2 (16.7)	1-4	20.0	0.0

COPD, Chronic Obstructive Pulmonary Disease

Items are presented in descending order of prevalence

Table 4.5. Convergent validity: Spearman's rank correlation coefficients of the self-reported number of conditions and DBMA with other health-related measurements

	Number of conditions	DBMA	p-level for correlation difference
Depression (CES-D) n= 474	0.35*	0.41*	0.0043
Physical functioning n= 596	-0.51*	-0.56*	0.0010
Quality of life (PWI) n= 540	-0.35*	-0.41*	0.0006
Perceived health n= 625	-0.51*	-0.56*	0.0035

*p <0.001. DBMA, Disease Burden Morbidity Assessment

4.2.2. FACTOR ANALYSIS

According to the screeplot (Figure 4.3), the exploratory factor analysis extracted 5 factors. Factor loadings >0.3 were considered to be included in the factors, although in the case of stroke, a slightly lower value was accepted due to its clear connection with one of the factors (Table 4.6). The following factors were found:

- Factor 1: conditions of the locomotor system (intermittent claudication, arthrosis, arthritis, osteoporosis and chronic back pain)
- Factor 2: depression/anxiety
- Factor 3: cardiovascular diseases (myocardial infarction, heart failure, angina, stroke)
- Factor 4: mixed group of cancer and renal/urinary tract diseases (kidney disease, cancer, urinary tract problems).
- Factor 5: lung disorders (asthma, COPD)

Five conditions did not fit in any of the factors: hypertension, diabetes, gastric/duodenal ulcer, memory disorders and Parkinson's disease. The explained variance was 43.6%.

Figure 4.3. Screeplot of the 21 DBMA items

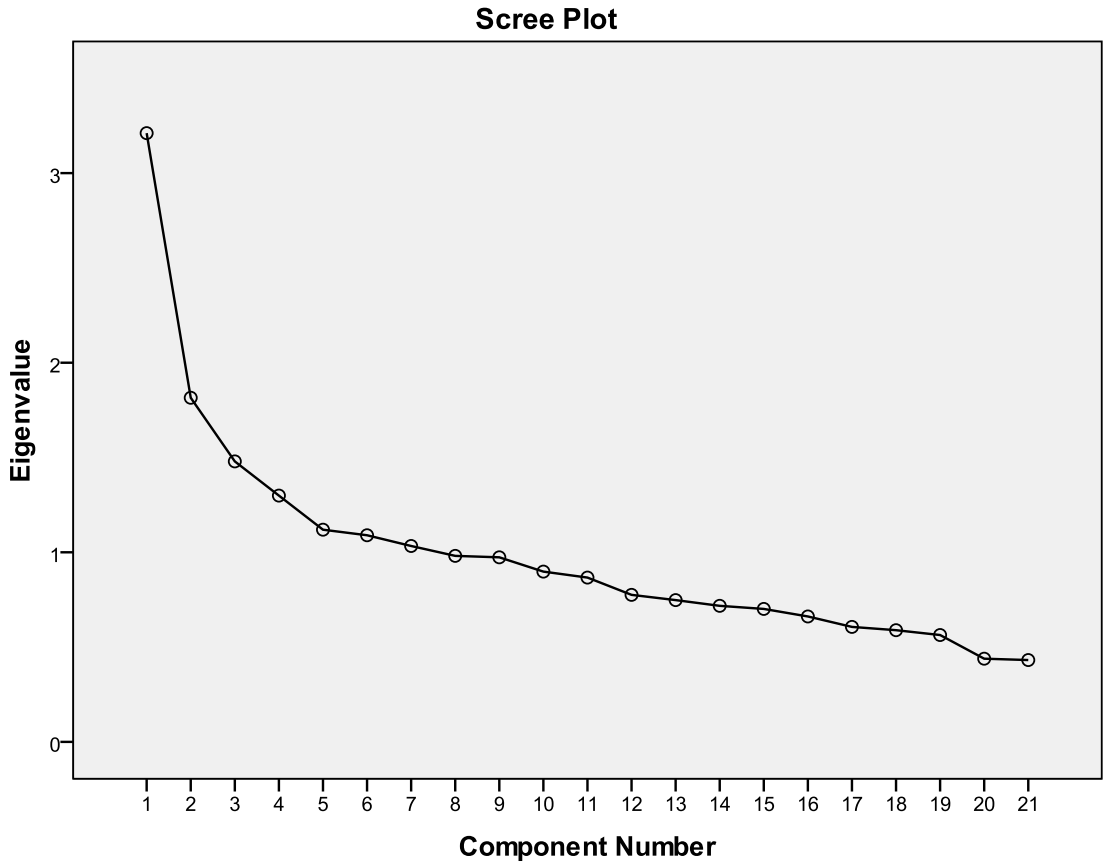


Table 4.6. Rotated factor loadings

Item		Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Uniqueness
1	Hypertension	0.05	0.25	0.11	0.16	0.15	0.88
2	Myocardial infarction	0.02	-0.08	**0.55**	-0.01	-0.05	0.69
3	Heart failure	0.04	0.00	**0.53**	0.01	0.01	0.71
4	Angina	0.00	0.10	**0.54**	0.00	0.16	0.68
5	Circulation problems/ intermittent claudication	**0.41**	0.20	0.15	0.10	0.04	0.75
6	Osteoarthritis	**0.65**	0.09	-0.01	-0.04	-0.04	0.57
7	Rheumatoid arthritis	**0.65**	0.10	0.07	-0.02	0.01	0.56
8	Asthma	0.01	0.27	0.04	0.06	**0.48**	0.69
9	COPD/emphysema	0.01	0.13	0.08	0.09	**0.50**	0.72
10	Diabetes	0.16	0.19	0.09	0.07	0.02	0.93
11	Gastric/duodenal ulcer	0.20	0.07	-0.01	0.08	-0.02	0.95
12	Kidney disease	0.05	0.16	-0.03	**0.43**	0.02	0.78
13	Depression	0.24	**0.64**	0.04	0.06	0.02	0.53
14	Anxiety	0.29	**0.59**	-0.04	-0.01	0.15	0.55
15	Cerebral embolism/stroke	0.05	0.15	**0.29**	0.03	-0.15	0.87
16	Cancer	0.06	-0.11	-0.03	**0.37**	0.04	0.84
17	Osteoporosis	**0.46**	0.16	-0.01	0.06	0.08	0.76
18	Memory disorders	0.07	0.04	-0.05	0.16	0.10	0.96
19	Parkinson's disease	0.11	0.06	-0.03	0.08	0.18	0.95
20	Chronic back pain	**0.54**	0.17	-0.08	0.18	-0.02	0.64
21	Urinary tract problems (prostate, bladder)	0.15	0.10	0.05	**0.47**	0.13	0.72

COPD, Chronic Obstructive Pulmonary Disease

Factor loadings in bold are those included in each of the factors: Factor 1, locomotor system; Factor 2, depression/anxiety; Factor 3, cardiovascular diseases; Factor 4, cancer and renal/urinary tract diseases; Factor 5, lung disorders.

4.3. RESULTS OF STUDY 2: CONSTRUCT VALIDITY

4.3.1. KNOWN-GROUPS VALIDITY BY SEX

Table 4.7.1 and 4.7.2 show the prevalences and mean disease burden scores per condition and by sex, for the total sample used in this study. Hypertension was the most frequent chronic health condition (n=297, 50.7%), and also the condition with the lowest mean disease burden score (mean score 1.6, SE=0.1). Parkinson's disease was the least frequent condition (n=8, 1.4%), and the condition with the highest mean disease burden score (3.0, SE=0.6), followed by chronic back pain (2.9, SE=0.2) and rheumatoid arthritis (2.9, SE=0.1).

Among men, hypertension was also the most frequent condition (n=130, 52.3%), but the lowest disease burden score was for gastric/duodenal ulcer (1.6, SE=1.3). Highest scores were found for asthma (3.0) (no SE provided in Stata because of a very low prevalence).

Among women, osteoarthritis was de condition with the highest prevalence (n=210, 64.8%). As in the total sample, hypertension was the condition with the lowest mean disease burden score (1.5, SE=0.1). The highest mean score was found for myocardial infarction (3.1).

Women had a mean number of chronic conditions of 3.76, men a mean number of 2.56 (p-value for the difference <0.001). Seven conditions were significantly more prevalent among women than in men: circulation problems, osteoarthritis, rheumatoid arthritis, depression, anxiety, osteoporosis, and chronic back pain. Five conditions were significantly more prevalent in men than in women: myocardial infarction, COPD, diabetes, cancer and urinary tract problems. For disease burden scores per present condition, five

significant differences were found, all with higher scores among women. This was the case for circulation problems, osteoarthritis, rheumatoid arthritis, diabetes and urinary tract problems.

Table 4.7. Prevalences and mean disease burden scores per condition and sex in the studied sample (n=625): men (n=280) vs. women (n=345)

		Prevalence†						Disease burden score						
		Total sample		Men		Women	Difference	Total sample		Men		Women		Difference
Item		Condition present		Condition present		Condition present								
		n	(%)	n	(%)	n	(%)	%	Mean	(SE) ‡	Mean	(SE) ‡	Mean	(SE) ‡	
1	Hypertension	297	(50.7)	130	(52.3)	167	(49.4)	-2.9	1.57	0.06	1.61	(0.09)	1.54	(0.08)	-0.07
2	Myocardial infarction	24	(4.1)	17	(6.1)	7	(2.5)	-3.6^	2.84	0.28	2.68	(0.35)	3.14	-	0.46
3	Heart failure	59	(9.7)	31	(10.4)	28	(9.2)	-1.2	2.42	0.20	2.04	(0.34)	2.77	(0.20)	0.73
4	Angina	27	(4.2)	12	(4.2)	15	(4.2)	0	2.06	0.22	1.84	(0.05)	2.23	(0.43)	0.39
5	Circulation problems/ intermittent claudication	137	(24.1)	37	(13.9)	100	(32.3)	18.4*	2.18	0.10	1.77	(0.14)	2.32	(0.13)	0.55**
6	Osteoarthritis	278	(47.1)	68	(25.1)	210	(64.8)	39.4*	2.77	0.08	2.45	(0.16)	2.87	(0.09)	0.42*
7	Rheumatoid arthritis	135	(22.3)	31	(9.7)	104	(32.5)	22.8*	2.91	0.13	2.51	(0.25)	3.01	(0.16)	0.52**
8	Asthma	33	(5.2)	15	(5.2)	18	(5.3)	0.1	2.60	0.17	2.99	-	2.30	(0.19)	0.31
9	COPD/emphysema	49	(8.4)	33	(12.1)	16	(5.4)	-6.7^	2.57	0.23	2.56	(0.11)	2.58	-	0.02
10	Diabetes	90	(15.6)	47	(17.6)	43	(13.9)	-3.7	1.94	0.12	1.64	(0.11)	2.23	(0.15)	0.59**
11	Gastric/duodenal ulcer	68	(11.3)	36	(13.0)	32	(10.0)	-3.0	1.76	0.16	1.56	(0.26)	1.98	(0.30)	0.42

COPD, Chronic Obstructive Pulmonary Disease; SE, Standard Error

† Unweighted counts and weighted percentages

‡ missing SE values are due to very low prevalences

*p<0.001; **p<0.01; ^p<0.05

Table 4.7. Prevalences and mean disease burden scores per condition per sex in de studied sample (n=625): men (n=280) vs. women (n=345) *continued*

		Prevalence†						Disease burden score							
		Total sample		Men		Women		Difference	Total sample		Men		Women		Difference
Item		Condition present n (%)		Condition present n (%)		Condition present n (%)			Mean	(SE) ‡	Mean	(SE) ‡	Mean	(SE) ‡	
12	Kidney disease	55	(9.4)	23	(8.8)	32	(10.0)	1.2	1.76	0.18	1.81	(1.34)	1.73	(0.23)	-0.08
13	Depression	109	(18.8)	20	(7.0)	89	(28.3)	21.3*	2.48	0.15	1.97	(0.55)	2.58	(0.15)	0.61
14	Anxiety	77	(13.1)	15	(5.6)	62	(19.2)	13.6*	2.60	0.16	2.44	(0.36)	2.63	(0.18)	0.19
15	Cerebral embolism/stroke	23	(4.4)	9	(3.6)	14	(5.0)	1.4	2.04	0.47	1.61	(0.44)	2.29	(0.98)	0.68
16	Cancer	67	(11.1)	42	(15.7)	25	(7.5)	-8.2**	2.03	0.20	1.88	(0.28)	2.29	(0.29)	0.41
17	Osteoporosis	101	(17.1)	6	(1.9)	95	(9.3)	7.4*	2.50	0.16	2.69	-	2.49	(0.16)	-0.20
18	Memory disorders	15	(2.7)	6	(2.3)	9	(2.9)	0.6	2.53	-	2.77	-	2.37	-	-0.40
19	Parkinson's disease	8	(1.4)	3	(1.0)	5	(2.0)	1.0	2.96	0.55	2.85	-	3.00	(0.60)	0.15
20	Chronic back pain	128	(22.3)	31	(11.6)	97	(30.9)	19.3*	2.92	0.15	2.55	(0.25)	3.03	(0.16)	0.48
21	Urinary tract problems (prostate, bladder)	113	(19.3)	78	(28.8)	35	(11.7)	-17.1*	2.21	0.13	1.95	(0.18)	2.73	(0.20)	0.78**

SE, Standard Error

† Unweighted counts and weighted percentages

‡ missing SE values are due to very low prevalences

*p<0.001; **p<0.01; ^p<0.05

4.3.2. KNOWN-GROUPS VALIDITY BY AGE GROUPS

Table 4.8 shows the prevalences and mean disease burden scores per condition, by age group (<75 years vs. ≥75 years). Hypertension was, as in the total sample, the most frequent condition in both age groups (n=155, 47.2% in the <75 years age group and n=142, 55.0% in the ≥75 years age group). Also, in both groups hypertension showed the lowest mean disease burden score, with a mean value of 1.6 (SE=0.1). In the <75 years age group, myocardial infarction (3.0, SE=0.2) showed the highest mean disease burden score. In the ≥75 years age group, the highest score was found for rheumatoid arthritis (3.0, SE=0.2).

Persons aged 75 years and older had a mean number of chronic conditions of 3.49, vs. 3.01 in persons aged <75 years (difference: 0.48, p=0.002). When studying these differences for single diseases, three conditions were significantly more frequent in the older group: rheumatoid arthritis, kidney disease and urinary tract problems. For one of the conditions (osteoarthritis) a significantly higher disease burden score was found among persons aged 75 years and older than among younger participants.

Table 4.8. Prevalences and mean disease burden scores per condition and age group in the studied sample (n=625): <75 years (n=350) vs. ≥75 years (n=275)

		Prevalence†					Disease burden score				
		<75 years		≥75 years		Difference	<75 years		≥75 years		Difference
	Item	Condition present n	(%)	Condition present n	(%)	%	Mean	(SE) ‡	Mean	(SE) ‡	
1	Hypertension	155	(47.2)	142	(55.0)	7.8	1.57	(0.09)	1.57	(0.09)	0
2	Myocardial infarction	10	(3.0)	14	(5.5)	2.5	2.98	(0.22)	2.74	(0.61)	-0.24
3	Heart failure	26	(7.8)	33	(12.1)	4.3	2.52	(0.27)	2.34	(0.30)	-0.18
4	Angina	14	(4.1)	13	(4.3)	0.2	2.22	(0.91)	1.86	-	-0.36
5	Circulation problems	64	(20.9)	73	(28.0)	7.1	2.15	(0.16)	2.21	(0.17)	0.06
6	Osteoarthritis	145	(44.5)	133	(50.2)	5.7	2.61	(0.11)	2.94	(0.13)	0.33*
7	Rheumatoid arthritis	60	(17.6)	75	(28.2)	10.6**	2.78	(0.16)	3.00	(0.20)	0.22
8	Asthma	18	(5.3)	15	(5.2)	-0.1	2.30	-	2.99	(0.17)	0.69
9	COPD/emphysema	24	(8.0)	25	(8.9)	0.9	2.36	(0.50)	2.80	(0.24)	0.44
10	Diabetes	46	(14.7)	44	(16.6)	1.9	2.17	(0.20)	1.67	(0.18)	-0.50
11	Gastric/duodenal ulcer	43	(13.0)	25	(9.3)	-3.7	1.76	(0.28)	1.78	(0.28)	0.02
12	Kidney disease	23	(7.0)	32	(12.4)	5.4**	1.68	(0.31)	1.83	(0.27)	0.15
13	Depression	61	(19.5)	48	(18.0)	-1.5	2.47	(0.24)	2.49	(0.18)	0.02
14	Anxiety	47	(15.4)	30	(10.4)	-5.0	2.53	(0.17)	2.72	(0.24)	0.19
15	Cerebral embolism/stroke	11	(3.6)	12	(5.3)	1.7	2.57	(0.71)	1.58	(0.31)	-0.99
16	Cancer	34	(10.7)	33	(11.7)	1.0	1.99	(0.40)	2.07	(0.32)	0.08
17	Osteoporosis	86	(17.4)	45	(16.7)	-0.7	2.41	(0.15)	2.61	(0.28)	0.20
18	Memory disorders	8	(2.6)	7	(2.7)	0.1	2.84	-	2.16	-	-0.68
19	Parkinson's disease	4	(1.1)	4	(1.8)	0.7	2.58	-	3.25	(0.43)	0.67
20	Chronic back pain	70	(22.1)	58	(22.5)	0.4	2.86	(0.18)	2.98	(0.24)	0.12
21	Urinary tract problems	45	(15.3)	68	(24.3)	9.0*	2.13	(0.28)	2.28	(0.22)	0.15

COPD, Chronic Obstructive Pulmonary Disease; SE, Standard Error

† Unweighted counts and weighted percentages

‡ missing SE values are due to very low prevalences

*p<0.01; **p<0.05

4.3.3. CONVERGENT VALIDITY: ASSOCIATION WITH PATIENT-CENTERED MEASURES AND HEALTHCARE UTILIZATION

All independent variables were significantly associated with the DBMA in the bivariate models (Table 4.9). The final multivariate regression model included 503 individuals, due to missing values in the independent variables. Female sex , primary/outpatient care and hospital care showed a significant and positive association with the DBMA. Perceived health, functional status, QoL, and affect balance and were negatively associated with disease burden. No significant associations were found for age. The model explained 42% of the variance.

Age was significantly associated with the DBMA in the bivariate analysis and no longer in the multivariate analysis. Therefore, a confounding analysis was performed (Table 4.10): perceived health and physical functioning were the variables that influenced the association with the DBMA in a negative way and showed significant associations with age.

Table 4.9. Association of patient-centered variables and healthcare utilization with disease burden: raw and adjusted linear models

	Bivariate analysis		Multivariate model	
Variable	Coefficient	p-value	Coefficient	p-value
Intercept			3.126	<0.001
Female sex	0.598	<0.001	0.192	0.024
Age	0.017	0.019	0.009	0.145
Very good/good perceived health	-0.928	<0.001	-0.424	<0.001
Physical functioning (no disability)	-0.933	<0.001	-0.532	<0.001
QoL (PWI)	-0.040	<0.001	-0.024	<0.001
Affect balance (SPANE)	-0.045	<0.001	-0.013	0.022
Primary/outpatient care in the past month	0.670	<0.001	0.358	<0.001
Hospital care in the past year	0.401	<0.001	0.186	0.040

R^2 for multivariate model=0.42, n=503

PWI, Personal Wellbeing Index; QoL, Quality of Life; SPANE, Scale of Positive and Negative Experience

Table 4.10. Regression coefficients between age and disease burden adjusted for each of the potential confounding variables separately and the relation between age and these variables

Regression coefficient			
Raw	0.017	Linear regression with age	
Adjusted for:		coefficient	p-value
Sex	0.017	0.45	0.415
Perceived health	**0.012**	**-1.57**	**0.006**
Physical functioning	**0.006**	**-3.44**	**<0.001**
Quality of life (PWI)	0.022	-0.04	0.170
Affect balance (SPANE)	0.018	0.02	0.625
Use of outpatient care	0.015	0.59	0.372
Use of hospital care	0.018	-0.61	0.300

PWI, Personal Wellbeing Index; QoL, Quality of Life; SPANE, Scale of Positive and Negative Experience

Variables in bold are those with a change-in-estimate of ≥20% and a significant association with age

4.3.4. PREDICTIVE VALIDITY: ASSOCIATION WITH MORTALITY

After four years, 35 of the 625 participants had died (5.5%). The Cox regression model showed a higher adjusted risk of death for persons with a higher score on the DBMA scale than for persons with lower scores: HR=1.073, 95% CI=1.002-1.148, p=0.044 (Table 4.11). When not correcting for complex samples, the HR remained practically the same (HR= 1.076, CI=1.027-1.128), but the p-value decreased to 0.002. The area under the ROC curve for the logistic regression model was 0.803 (Figure 4.4), with a 95% CI of 0.727-0.879.

Figure 4.5 shows the Kaplan-Meier curves for different DBMA categories. Persons with DBMA scores ≤ 2 had a higher survival rate than persons with higher DBMA scores. Although less pronounced, the remaining two categories (3-10, ≥11 DBMA scores) followed the same trend, with the lowest survival rate for persons in the highest DBMA-score category.

Table 4.11. Disease burden as a predictor of mortality: Cox regression

	Corrected for complex samples (n=503)				Raw data (n=625)			
Variable	HR	SE	p-value	95% CI	HR	SE	p-value	95% CI
DBMA	1.073	0.037	0.044	1.002-1.148	1.076	0.026	0.002	1.027-1.128
Age	1.085	0.035	0.013	1.017-1.157	1.108	0.029	<0.001	1.051-1.167
Sex	0.108	0.082	0.004	0.024-0.483	0.099	0.048	<0.001	0.038-0.257

CI, Confidence Interval; DBMA, Disease Burden Morbidity Assessment; HR, Hazard Ratio; SE, Standard Error

Figure 4.4. Receiver Operating Characteristic (ROC) curve of the logistic regression model with mortality as a dependent variable

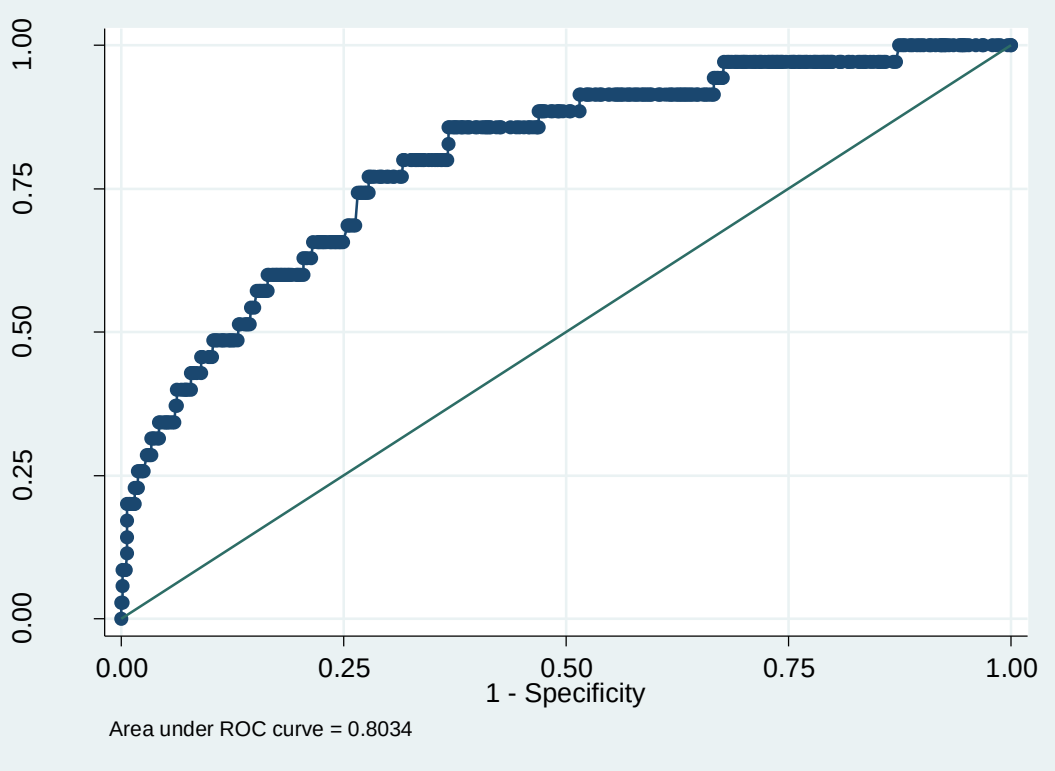
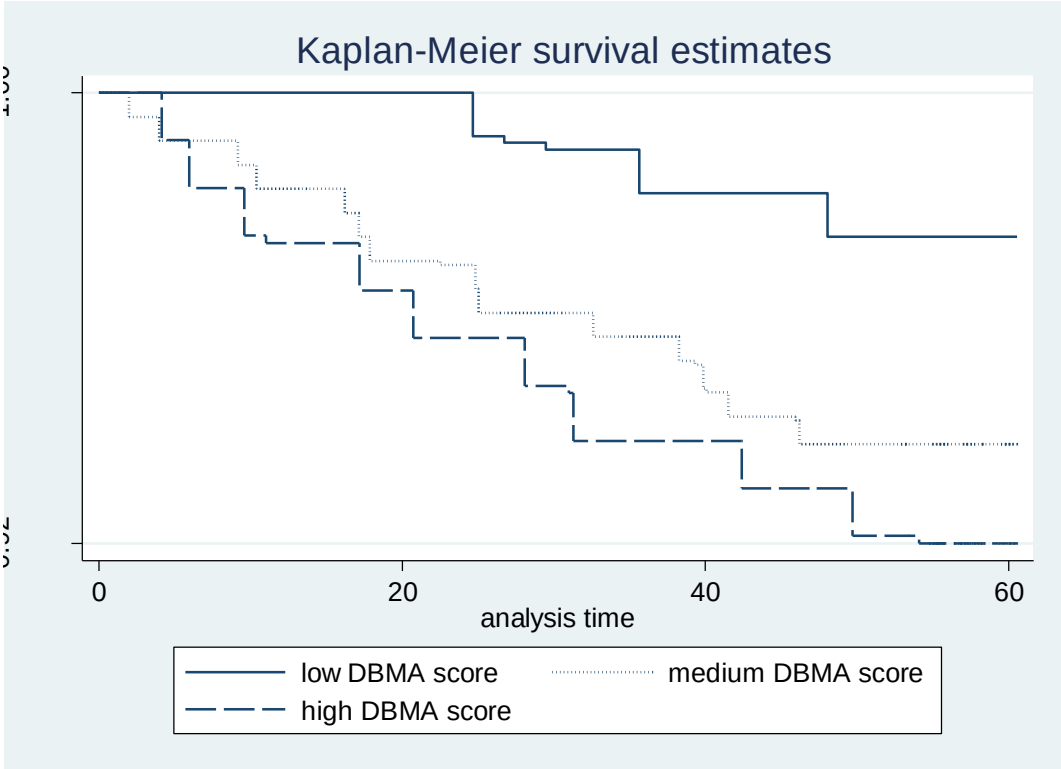


Figure 4.5. Kaplan Meier probability of survival per DBMA category



DBMA, Disease Burden Morbidity Assessment

Applied DBMA cut-off points: low, 0-2; medium, 3-10; high, ≥11.

4.4. RESULTS OF STUDY 3: RASCH ANALYSIS

4.4.1. RASCH ANALYSIS

The initial analysis, with the whole study sample (n=1400), displayed poor fit to the Rasch model (Table 4.12). After selecting a subsample of 300, the fit indices improved, but still did not meet the fit criteria. Category probability curves showed disordered thresholds, so items were rescored to two (two items), three (13 items), four (five items) or four categories (one item) (Table 4.13). After this, the DBMA showed an acceptable fit to the Rasch model (Table 4.12). Individual item and person fit residuals were within the -2.5 to +2.5 range, with non-significant chi-squares (Table 4.13). However, PSI remained low, 0.272. When repeating this estimation in RUMM2020 the PSI improved to 0.637.

In the PCA of the residuals, 0.72% of tests were outside the previously set range, indicating unidimensionality. All items were locally independent, with a residual correlation index ranging 0.000-0.188. No DIF was found for age or educational level. Four items showed DIF by sex of small magnitude (< 0.5 logits): item 1 (hypertension), 14 (anxiety), 17 (osteoporosis) and 21 (urinary tract problems). In the top-down purification approach this DIF was no longer present. The person-item threshold distribution (Figure 4.6) showed a floor effect and no persons represented the scale's higher levels of disease burden.

DBMA scores of the total sample were converted into a linear measure from 0 to 47 (Table 4.14). When comparing the subsample of 300 and the rest of the sample, the differences between the estimations were not significant (difference=0.259 logits, t-test= 1.226, p-value=0.226), and there was no DIF by sample.

Table 4.12. Global fit to the Rasch model of the DBMA using the total sample (n=1400), after selecting a subsample (n=300) and after rescoring the response scale

		Standard	Total sample	Subsample	After rescoring
Item fit residual	Mean	0	-2.90	-1.12	-0.68
	SD	1	1.83	0.77	0.71
Person fit residual	Mean	0	-0.48	-0.42	-0.28
	SD	1	0.59	0.58	0.55
Item-trait interaction	χ^2	Low	316.33	165.14	154.43
	Prob.	NS	<0.001	0.89	0.32
PSI		>0.70	0.07	0.14	0.27
Unidimensionality	Significant t-tests (%)	< 5	0.72	2.00	2.00

NS, non-significant; PSI, Person Separation Index; Prob, Probability; SD, Standard Deviation

Note: Item fit residual refers to the difference between the data observed and the expected values at item level.
Person fit refers to the difference between the data observed and the expected values at person level.
Item-trait interaction is a chi-square value and probability resulting from the comparison between the expected and the mean observed score for groups of people with similar ability estimates.
Person separation index is a reliability measure.
Unidimensionality refers to the existence of one measurement construct (dimension) underlying the set of items.

Table 4.13. Threshold ordering of polytomous items and individual item fit to the Rasch model after rescoring (n=300)

	Original categories						Individual item fit to the Rasch model				
	0	1	2	3	4	5					
Item	Rescored categories						Location	SE	Residual	χ^2	Prob.
Osteoarthritis	0	1	1	1	2	3	-1.914	0.093	-2.020	14.737	0.040
Rheumatoid arthritis	0	1	1	1	2	3	-1.514	0.108	-1.749	5.988	0.541
Chronic back pain	0	1	1	1	1	2	-1.392	0.130	-0.707	10.910	0.143
Depression	0	1	1	1	1	2	-1.161	0.140	-0.996	7.631	0.366
Circulation problems/ intermittent claudication	0	1	1	1	2	3	-0.965	0.134	-1.698	5.237	0.631
Hypertension	0	1	2	2	3	4	-0.934	0.088	-0.515	16.122	0.024
Anxiety	0	1	1	1	1	2	-0.858	0.160	-1.474	7.290	0.399
Osteoporosis	0	1	1	1	1	2	-0.644	0.162	-1.271	3.796	0.803
Cancer	0	1	1	1	2	3	-0.515	0.160	0.068	8.254	0.311
Diabetes	0	1	1	1	1	2	-0.371	0.174	-0.637	6.877	0.442
Heart failure	0	1	1	1	1	2	-0.331	0.196	-0.055	8.192	0.316
Urinary tract problems (prostate, bladder)	0	1	1	1	2	3	-0.219	0.147	-0.436	6.327	0.502
COPD/emphysema	0	1	1	1	1	2	-0.137	0.222	-0.162	5.471	0.603
Cerebral embolism/stroke	0	1	1	1	1	1	0.296	0.353	0.171	10.195	0.178
Memory disorders	0	1	1	1	1	1	0.310	0.355	-0.424	5.818	0.561
Gastric/duodenal ulcer	0	1	1	1	1	2	1.081	0.199	0.849	9.992	0.189
Kidney disease	0	1	1	1	1	2	1.190	0.218	-0.444	7.298	0.399
Asthma	0	1	1	1	1	2	1.578	0.274	-1.258	5.272	0.627
Myocardial infarction	0	1	1	1	1	2	1.685	0.296	-0.340	2.781	0.904
Angina	0	1	1	1	1	2	2.039	0.473	-0.312	4.173	0.760
Parkinson's disease	0	1	1	2	2	2	2.776	0.924	-0.912	2.071	0.956

COPD, Chronic Obstructive Pulmonary Disease; SE, Standard Error; Prob., Probability

Items are ordered by increasing difficulty (mean location of thresholds)

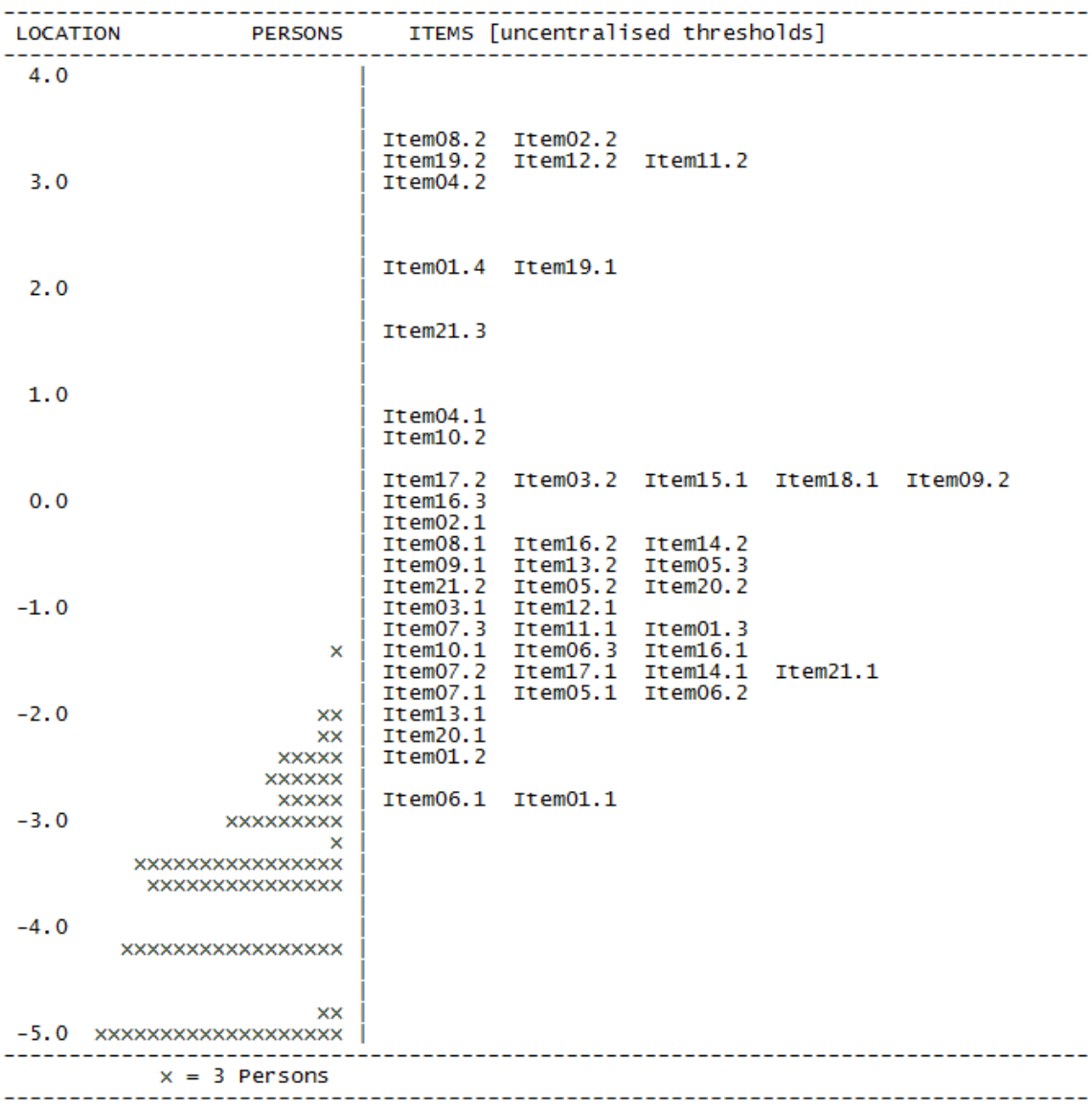
Table 4.14. Initial score to linear measure conversion table (final Rasch model), n=1400

Initial score	Linear measure (logit scale)	Linear measure (0-47)	Initial score	Linear measure (logit scale)	Linear measure (0-47)
0	-4.178	0.000	24	-0.164	23.202
1	-3.406	4.462	25	-0.085	23.659
2	-2.891	7.439	26	-0.005	24.121
3	-2.547	9.428	27	0.076	24.590
4	-2.287	10.931	28	0.158	25.064
5	-2.075	12.156	29	0.241	25.543
6	-1.897	13.185	30	0.326	26.035
7	-1.741	14.087	31	0.413	26.538
8	-1.602	14.890	32	0.503	27.058
9	-1.477	15.613	33	0.596	27.595
10	-1.362	16.277	34	0.693	28.156
11	-1.255	16.896	35	0.794	28.740
12	-1.155	17.474	36	0.901	29.358
13	-1.060	18.023	37	1.014	30.012
14	-0.969	18.549	38	1.136	30.717
15	-0.882	19.052	39	1.268	31.480
16	-0.798	19.538	40	1.415	32.329
17	-0.715	20.017	41	1.578	33.272
18	-0.635	20.480	42	1.766	34.358
19	-0.555	20.942	43	1.988	35.642
20	-0.477	21.393	44	2.261	37.220
21	-0.398	21.850	45	2.620	39.295
22	-0.320	22.301	46	3.156	42.393
23	-0.242	22.751	47	3.953	47.000

Note: This table is not applicable when there are missing data.

To use this table, first score items according to Table 4.13. The sum of items scores is the "initial score", which can be converted to a linear measure in logits (second column) or a 0-47 scale (third column).

Figure 4.6.Person-item threshold distribution map: final Rasch analysis of the DBMA



Item01. Hypertension; Item02. Myocardial infarction; Item03. Heart failure; Item04. Angina; Item05. Circulation problems/ intermittent claudication; Item06. Osteoarthritis; Item07. Rheumatoid arthritis; Item08. Asthma; Item09. COPD/emphysema; Item10. Gastric/duodenal ulcer; Item11. Gastric/duodenal ulcer; Item12. Kidney disease; Item13. Depression; Item14. Anxiety; Item15. Cerebral embolism/stroke; Item16. Cancer; Item17. Osteoporosis; Item18. Memory disorders; Item19. Parkinson's disease; Item20. Chronic back pain; Item21. Urinary tract problems (prostate, bladder)

Note: Persons are represented on the left side, each 'x' representing 3 persons. On the right side of the graphic, item thresholds are represented, with the corresponding threshold for each item after the dot.

4.4.2. CLASSIC PSYCHOMETRIC ANALYSIS OF THE LINEAR MEASURE

The distribution of the linear measure is shown in Figure 4.7. Mean score was 7.36 (SD=5.01), median score 7.44, with a mean-median difference of 0.17%. Floor effect for the total scale was 18.11%, with no ceiling effect, and skewness was 0.046. The linear measure presented a correlation of -0.48 with physical functioning, -0.47 with perceived health, 0.32 with depression (CES-D) and -0.24 with PWI ($p<0.001$). Women scored significantly higher than men, with mean scores of 8.14 (SD=5.15) and 6.40 (SD=4.65), respectively ($p<0.001$), and scores increased with age: mean score among persons <65 years was 5.93 (SD=4.76) vs. a mean score of 8.77 (SD=4.84) in persons aged ≥65 years ($p<0.001$).

The results of the relative precision analysis are shown in Table 4.15. The ability to discriminate between age groups increased by 9% when using the linear measure vs. the raw score (95%CI: 1.03-1.17), but precision decreased 4% for age groups, although this difference was not statistically significant (95% CI: 0.86-1.05).

Figure 4.7. Distribution of the linear measure (n=1397)

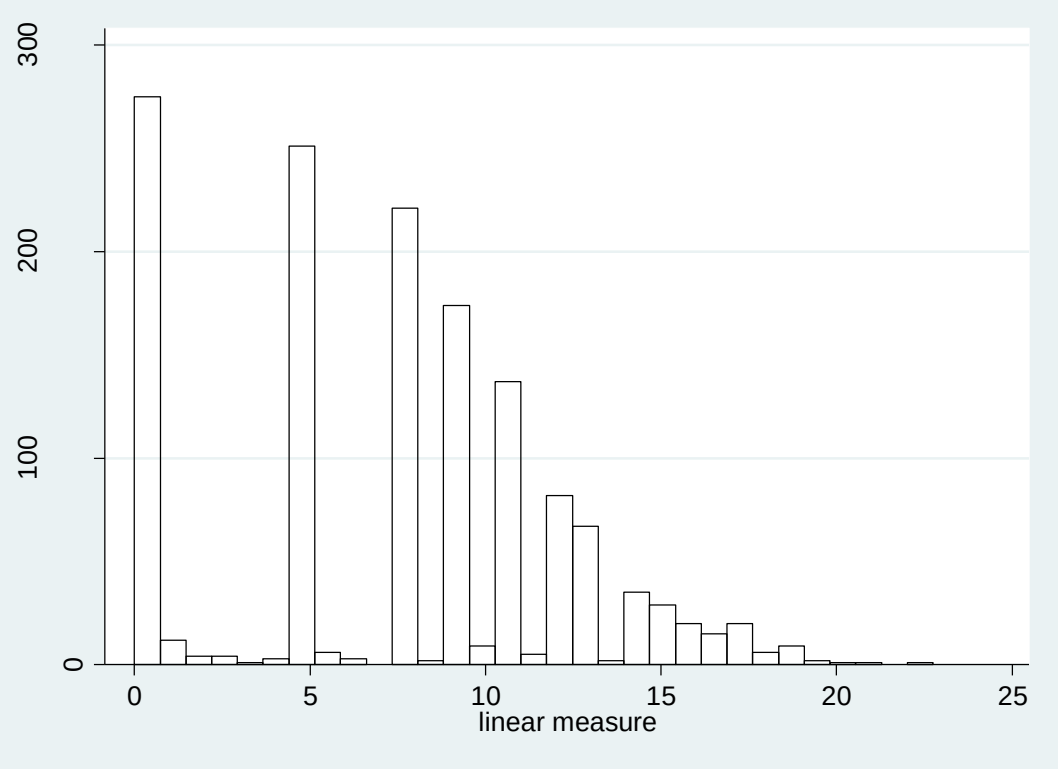


Table 4.15. Relative precision of the linear measure in comparison to the raw summative DBMA score (n=1277)

Scoring method	Patient groups	Mean (SE)	Mean difference (SE)	Z statistic	RP	95% CI
Raw score	<65 years	3.81 (0.20)	3.02 (0.35)	-9.89	1.00	
	≥65 years	6.83 (0.28)				
Linear measure	<65 years	5.95 (0.19)	2.98 (0.26)	-10.77	1.09	1.03-1.17
	≥65 years	8.92 (0.19)				
Raw score	Men	3.88 (0.19)	2.59 (0.35)	-6.61	1.00	
	Women	6.47 (0.28)				
Linear measure	Men	6.43 (0.19)	1.78 (0.27)	-6.33	0.96	0.86-1.05
	Women	8.21 (0.19)				

CI, Confidence Interval; SE, Standard Error; RP, Relative Precision

5. DISCUSSION

This work describes validation of the DBMA, a comorbidity assessment instrument that incorporates a dimension not usual in other comorbidity measures: the disease burden caused by the present chronic conditions. Co- or multimorbidity requires a patient-centered healthcare approach, with a more holistic view that goes beyond treating single diseases (8). Since chronic conditions usually cannot be cured, the goal of caring for persons with multimorbidity should be mostly centered in maximizing QoL (41). Therefore, interventions to improve care for this population often assess subjective outcomes as QoL and emotional wellbeing rather than objective outcomes. These subjective outcomes reflect what is most meaningful to the patients themselves. However, patient reported outcomes may be influenced by other factors than multimorbidity itself. The DBMA provides a subjective measure that is directly related to multimorbidity. By asking patients about the impact of chronic conditions on their daily activities, without specifying which activities or the kind of impact, the DBMA measures not only physical but also psychological limitations on a broad spectrum of everyday activities. This makes the DBMA an extremely useful subjective measure of multimorbidity, or better said, the disease burden caused by multimorbidity.

This thesis presents the results of a validation of the DBMA according to the CTT (objective 1: feasibility, acceptability, scaling assumptions, reliability, construct validity and exploratory factor analysis), including known-groups, convergent and predictive validity (objective 2), and Rasch analysis (objective 3: test of fit to the Rasch model, reliability, unidimensionality, response dependency, category structure, DIF, scale targeting and CTT analysis of the linear measure).

5.1. STUDY 1: VALIDATION OF THE DBMA ACCORDING TO THE CTT

In order to perform a validation according to the CTT, the following psychometric properties were studied: feasibility, acceptability, scaling assumptions, reliability and construct validity. Dimensionality was assessed through an exploratory factor analysis.

Feasibility was assessed by determining the percentage of missing values per item and the percentage of computable scores for the total scale. The percentages of missing values per item were all below the 10% criterion. However, the percentage of computable scores was just below the criterion of 90%. The DBMA in this study was part of the CAPI questionnaire which consisted of 218 questions. The fact that the DBMA was included in such a large questionnaire might have influenced the willingness and capability to answer the questions of the DBMA.

Acceptability was explored by comparing possible and observed scores, mean-to-median difference for the total scale as well as floor and ceiling effects and skewness. The observed scores covered the complete response ranges for the 21 items, except for Parkinson´s disease, which had a response rate of 0-4 but also a very low prevalence. For the total scale a response range of 0-41 was observed, which was way below the hypothetic maximum score of the DBMA of 105. However, the DBMA was not designed to cover the complete total score range, since this would mean the presence of all 21 chronic conditions in the same person, with a maximum impact on daily life. Therefore, the observed range coincided with our expectations. Mean to median difference was within the set criterion.

One of the most remarkable findings in this analysis has been the floor effect (and, as a consequence, the high value for skewness). The high floor effects represent cases in which

participants reported not having certain conditions. This implies that the floor effect can actually be regarded as an indicator of how 'healthy' the studied population is. The sample used in this study consisted of community-dwelling adults, with quite a high health status. Institutionalized persons were not included, which made that the ELES-PS sample was healthier than the general population. Rodríguez Laso et al. (105) conducted a study about selection bias in the ELES-PS. They found that the persons that refused to participate in the CAPI interview, the questionnaire that contained the DBMA, were of higher age and reported lower perceived health than the respondents who did answer this questionnaire (105). Also, only persons with household telephone lines were selected. The proportion of persons aged 50 years and older in Spain with telephone lines is estimated to be at least 92% (105), but it is possible that people that do not have a telephone line have lower incomes, which is known to be related to lower health status (106). Nevertheless, even when using the DBMA in a sample with high prevalence of multimorbidity, high floor effects per item would still be found, because there would still be a high number of non-present conditions in every participant.

When studying disease burden scores for only the present conditions, the floor effects persisted, though less pronounced. This means that the participants in the used sample often responded that they did not experienced any limitation because of their chronic conditions. Hypertension, kidney disease and gastric/duodenal ulcer all showed floor effects above 60%. In the case of kidney disease, this is a little surprising. Chronic kidney diseases are known to be associated with poor physical function and to have a negative effect on activities of daily living (107–109). This raises the question whether the participants really did not experience disease burden from their kidney diseases or if they might have wrongly qualified any other conditions as kidney diseases. Including examples of the listed conditions, as was done by Bayliss et al (41), could help to prevent confusion about the listed conditions. Another option would be to ask the participants to specify

which particular conditions they have, however, this would make the questionnaire more laborious, both for the respondent and for the interpretation of the questionnaires.

Two conditions showed a ceiling effect when only including present conditions: COPD/emphysema and memory disorders. This means that these were the two conditions with the highest proportion of a maximum disease burden. However, also these two conditions showed important floor effects. It seems like for these two conditions, the limitations followed an 'all nor nothing' pattern: either they do not limit daily activities, or they do a lot. In any case, the high ceiling effects coincide with the literature, as both conditions are known to cause important limitations of daily living activities. In case of COPD these limitations are mainly physical (110–112); for dementia it is the cognitive impairment that might limit patients in their daily life (113,114).

One more comment should be made about the floor- and ceiling effects. As can be seen in Table 3.1, and as done by Bayliss et al. (45), the middle response categories of the response scale were not labelled. This could have prevented participants from choosing the middle options. Further research might include these labels in order to assess whether this would make participants more likely to choose the middle response options.

Scaling assumptions were determined through the ITCC for each item. There were only six conditions that met the criterion: osteoarthritis, intermittent claudication, rheumatoid arthritis, chronic back pain, depression and anxiety. These conditions have in common is that they are all among the 10 most prevalent conditions in our sample. Apparently, the ITCC improves when the prevalence increases. This implies that also the low ITCC for the remaining conditions could be explained by the fact that the ELES-PS sample was relatively healthy, and would probably improve in more comorbid populations. We did not expect to find high ITCC because the scores for individual conditions are not necessarily

related to the total scores: experiencing high disease burden from one disease does not mean that the other 20 conditions should be present as well. However, having comorbid conditions does increase the disease burden experienced from a specific disease (115,116), which might have caused the ITCC to be higher in the more prevalent conditions.

Reliability was assessed through Cronbach's alpha, which did meet the criterion for group comparisons. Cronbach´s alpha is a measure that indicates to what extent the items measure the same construct. In this case, the construct was disease burden, and Cronbach´s alpha confirmed that this was the same for all items. Reliability will be further addressed in the discussion section about Rasch analysis.

The item homogeneity index, the mean of inter-item correlations, was extremely low. As explained above, the items of the DBMA were not designed to be associated to each other nor to the total score, and therefore, it was not a surprise to find a low homogeneity index.

For convergent validity, we expected self-reported disease burden to be negatively associated with perceived health, physical functioning and QoL and to find a positive association with depression (10,80,81). Furthermore, these correlations were compared with those of the simple disease count of the 21 conditions included in the DBMA. ELES-PS data confirmed that the DBMA is more strongly associated with depression, perceived health and physical functioning than the number of conditions, although these differences were less pronounced than in the study published by Bayliss and colleagues (41). This outcome is of interest because, although the DBMA is a relatively short questionnaire, it is still more laborious than a simple disease count. The association between a simple count of chronic conditions as a measure of multimorbidity and QoL has widely been studied (117,118). However, measures taking into account disease severity seem to be better predictors of QoL (119), as was confirmed in this study. The relationship between the

DBMA and the PWI, a general QoL measure, had not been studied yet and this study has found a moderate correlation. Some of the dimensions of this scale are not directly related to health, such as personal safety and future security (73), and the correlation with a scale measuring health related QoL could possibly be stronger.

Known-groups validity was examined comparing disease burden by sex and age groups. Our hypothesis to find higher scores in women and older people was confirmed. The difference between men and women was especially pronounced. This can be explained by the fact that not only multimorbidity is more frequent among women, but women also experience more functional limitations for every additional chronic condition (120). This will be further addressed in the discussion section about Study 2.

The factor analysis identified known disease groups (121). Intermittent claudication is a motor symptom and therefore it was probably more related to musculoskeletal disorders than to cardiovascular diseases. Dimensionality was also assessed in Rasch analysis, so the findings of this factor analysis will be further discussed below.

In summary, Study 1 found satisfactory feasibility and acceptability for the DBMA, except for large floor effects, which could be explained by the design of the DBMA as well as by the high health status of the used sample. Items were not highly related to each other nor to the total score (low ITCC and item homogeneity index) but, according to Cronbach´s alpha, DBMA items do measure the same construct: disease burden. The construct validity analysis found the DBMA to be related to other patient-centered outcomes. Construct validity was further studied in Study 2.

5.2. STUDY 2: KNOWN-GROUPS, CONVERGENT AND PREDICTIVE VALIDITY

In this study, known-groups validity by sex and age groups was assessed more thoroughly. In addition convergent validity was further assessed by analyzing the association between the DBMA, other patient-centered outcomes and healthcare utilization. Finally, predictive validity was studied through the association with mortality.

5.2.1. KNOWN-GROUPS VALIDITY

In the known-groups validity analysis of Study 1, women and persons aged 75 years and older had significantly higher disease burden scores. In Study 2, this was assessed per condition. The mean number of present conditions and disease prevalences were also compared for these groups, in order to find out whether the results found in Study 1 were due to differences in disease prevalence, or due to differences in disease burden per present condition.

As described in the discussion of Study 1, we expected disease burden per condition to be higher in women, since women are known to experience more functional limitations per present condition (120). Indeed, for five conditions, women reported a significantly higher mean disease burden score per present condition. Most of the differences in disease burden we not significant, which can be ascribed to the low prevalences of the conditions, resulting in very few disease burden scores per condition. Significant differences were found for the most prevalent conditions. Women had a significantly higher mean number of present conditions than men and seven of the conditions were more frequent in women, versus five conditions that were more prevalent in men. From these data, it can be

concluded that the women in this sample showed higher disease burden scores, caused by both higher disease prevalences and higher disease burden per condition.

The question remains why women experience higher disease burden per present condition than men. This question cannot be easily answered. Do women have more serious diseases or do they 'complain' more? A possible explanation, as described by Hunt et.al.(122), is that women have more traditional "role obligations" than men, since their duties include more elements of caretaking. This article was published in 1984 and more than 30 years have gone by since then, but these roles probably remain largely unchanged among older Spanish adults.

The increase in disease burden with age was less clear. Only one condition, osteoarthritis, showed a significantly higher disease burden when present in the ≥75 years of age group. The mean number of conditions per person was higher in older participants and three conditions were significantly more frequent among the highest age group. These results suggest that for age, the differences in disease burden scores were more due to differences in disease prevalence rather than differences in the disease burden caused by every present condition. This conclusion coincides with a study performed by Henchoz et al (123). They found that, although the number of conditions increased rapidly with age in octogenarians, their perceived health decreased in a much less steep way. This was explained by the fact that older people compare their health to that of persons of the same age. Since chronic conditions are frequent among older people, participants often concluded that their own health status was not that bad or it was even better than their peers' health. The same mechanism might be applied to disease burden. Older persons might be more limited by their conditions, but since limitations are common in their age group, they do not experience them as serious as younger people.

5.2.2. CONVERGENT VALIDITY

In the multivariate analysis, significant associations between the DBMA and perceived health, functional status, QoL, and affect balance were found. These results confirm the associations found in previous studies between the DBMA and patient-centered outcomes, (41,45,46) and are congruent with findings from Study 1. The relation between the DBMA and the SPANE as a measure of affect balance had not been studied before. A negative association between affect balance and self-reported disease burden was found. This was expected, since studies have shown that the SPANE is negatively associated with depression (87), and depression positively associated with the DBMA (41).

The multivariate regression showed no significant relation between the DBMA and age, but female sex was positively associated with self-reported disease burden, indicating that women had higher scores on this scale, as was also concluded in Study 1 and the known-groups validity assessment in the current study. When assessing confounding factors, perceived health and physical functioning had a negative impact on the association between age and the DBMA. These findings are consistent with the literature since older people have lower perceived health(124) and poorer physical functioning(125), and these variables themselves are associated with the DBMA.

The use of healthcare resources was added in this analysis as a system-centered outcome. The term system-centered outcome can be described as an outcome that can be expressed in healthcare costs, whereas patient-centered outcomes are variables that are of high importance for individual patients, such as QoL (126).The measures of healthcare use were self-reported, which implies that some level of subjectivity should be taken into account. In a recent Dutch study, community-dwelling older persons slightly overestimated healthcare utilization, and this was more frequent among people with

multimorbidity and among those who reported having a worse heath-status than a year before (127). However, their overall conclusion was that self-report of healthcare utilization in older community-dwelling persons was adequate and efficient.

A significant positive association between the DBMA and primary/outpatient care and hospital care was found. This second association was weaker and less significant. This can be due to two reasons. First, primary/outpatient care was twice as frequent in the used sample as hospital care. Second, the hospital care variable referred to the past year, whereas primary/outpatient care referred to month prior to the interview, which could make it more related to the disease burden experienced at that moment.

Bayliss et al. previously studied the relationship between the DBMA and utilization outcomes (46). They reported significant associations with outpatient utilization and inpatient admissions; the relation with emergency department admissions was not significant. They did the same analysis for the Charlson Comorbidity Index (CCI), finding significant associations with the three utilization outcomes with stronger associations than the DBMA. Their results found for the DBMA are similar to the results in the current study; it seems that there is a relationship between the DBMA as a patient-centered multimorbidity scale and the system-centered outcome of healthcare utilization. However, healthcare utilization was more related to the CCI and it might be more appropriate to use a more system-centered multimorbidity scale such as the CCI when predicting the use of healthcare resources (26,46).

Descriptive statistics were presented for the whole sample used in Study 2 as well as for the sample used in the multivariate regression model. Due to missing values in the variables in the multivariate regression model, this sample was reduced to 503. However,

the reduced sample resulted quite similar to the total sample, so this reduction should not have influenced the conclusions drawn in this analysis.

5.2.3. PREDICTIVE VALIDITY

For predictive validity, the longitudinal association of the DBMA with four-year mortality was assessed. The DBMA was designed and validated to be associated with patient-centered outcomes, unlike other multimorbidity assessment scales such as the CCI and Elixhauser´s comorbidity measure where mortality was one of the main outcomes of interest (24,29). To the best of our knowledge, the relationship between the DBMA and mortality had not been studied yet. A positive association between the DBMA and mortality after four years follow-up time was found. For every point of increase of the DBMA total score, there is a 7 % increased risk of death after four years. The Cox analysis was repeated without taking into account the correction for complex samples, because of the few deaths that took place in the ELES-PS 'healthy' community-dwelling population. After correction, the HR remained practically the same, but the p-value, which was on the border of significance, decreased to 0.002. This confirms that the higher p-value of 0.044 was because of the decreased sample size that resulted from the correction for complex samples.

The Kaplan-Meier curve showed a higher mortality rate among the highest DBMA categories. The difference between low DBMA scores (0-2) and the rest of score categories was especially pronounced. These findings support the association found in the Cox analysis. Finally, the area under the ROC was 0.803, which implies that the model can be regarded as moderately predictive (89). This confirms that the DBMA may be used as a predictor of mortality, and adds information about the applicability of the DBMA.

In this study, apart from descriptive statistics of ELES-PS data, data from the Spanish Population and Housing Census and the Spanish National Health Survey were presented, as well as national mortality rates. The participants in our sample showed higher

perceived health and the mortality rate found among its participants was way lower than the expected. This confirms once again that the ELES-PS sample was relatively healthy in comparison to the general population. Nonetheless, it is possible that the mortality rate was slightly underestimated because of matching errors between the ELES-PS database and the Ministry of Health's. The search was repeated manually in order to avoid this, but it is not guaranteed that no other case was missed.

It was not possible to compare the associations that were for found for the DBMA with those of another multimorbidity assessment instrument (concurrent validity), since no other multimorbidity measures were included in the ELES-PS survey. Comparisons with the CCI were performed before for the association with patient-centered and utilization outcomes, showing that the DBMA was more related to the former and less to the latter than the CCI (41,46), but it would be interesting to compare its performance predicting mortality. Further research should include this comparison with other multimorbidity tools which would facilitate the selection of the most appropriate instrument depending on the outcomes of interest.

Briefly, Study 2 found significant relations between the DBMA and patient-centered outcomes, the system-centered variable of healthcare utilization and mortality. Although not assessed in the current study, previous research showed the CCI to be less related to patient-centered outcomes and more related to healthcare utilization than the DBMA. This implies that when choosing a comorbidity assessment instrument, this should be done according to the outcome of interest. If the outcome of interest is patient-centered, such as QoL, results of the current and previous studies suggest that the DBMA would be a good option.

Women showed not only higher disease prevalences but also higher disease burden scores per condition. According to these findings, interventions to improve QoL among multimorbid patients should put a special focus on women. The DBMA might be applied to identify other groups of patients with high disease burden relative to their disease prevalences, such as groups according to income, educational level, or certain combinations of conditions.

5.3. STUDY 3: RASCH ANALYSIS

Study 3 analyzed the measurement properties of the DMBA according to the Rasch model. Rasch analysis provides knowledge of psychometric attributes that are not assessed with CTT: test of fit to the Rasch model, response dependency, category structure, DIF and scale targeting. Other attributes, namely reliability and unidimensionality are assessed in both approaches but in a different manner. Moreover, Rasch analysis provides a linear measure, which allows calculation of change scores and, given a normal distribution, the use of parametric statistics (57,128).

The Rasch analysis was performed with a randomly selected subsample of 300, since analysis with samples larger than 300 could result in statistically significant deviations from the Rasch model of otherwise well-fitting items. Linacre (64) stated about this:

If the test involves less than 30 observations, it is probably too insensitive, i.e., "everything fits". If there are more than 300 observations, it is probably too sensitive, i.e.,"everything misfits".

This was also explained in an article published by Smith et al. (66): some Rasch fit statistics for polytomous scales are highly dependent on sample size, which translates into a higher possibility for type I errors with an increased sample size. Other studies used subsamples for the same reason (65,129). In order to compare the subsample with the total sample, descriptive statistics of both samples were paralleled, as well as the logit estimations of the two samples (300 vs. 1100) for each raw-score. In addition, a DIF analysis with the sample as a factor was performed. According to these comparisons, the subsample did not significantly differ from the total sample, meaning that the subsample forms an unbiased representation of the total sample.

To achieve an adequate fit to the Rasch model, items needed to be rescored. This might have been due to too many response categories, which could have prevented people from making fine distinctions between rating scale steps. Another possible cause is the fact that the middle response options were not labeled. In most cases, response options were reduced to three categories. This reduction in response categories does not require changing the original questionnaire. Instead, it may be performed when calculating the total scores, thus avoiding using different response categories that could be confusing to the respondent. It would be interesting to study whether simplifying the questionnaire, by reducing the response categories in the same way for all items, would improve the psychometric properties of the DBMA. However, this might reduce the scale precision. Labeling the middle response options may also improve category structure. Further research is needed to confirm the response structure.

The residual correlation index did not identify response dependency, meaning that there were no items linked in such way that the response to one item would determine the response to another. The residual correlations of the items of the DBMA were all far below the cut-off point, so no items needed to be combined.

Four items displayed DIF by sex, indicating that men and women, despite having the same level of burden caused by hypertension, anxiety, osteoporosis or urinary tract problems, answered differently to these items. A very strict approach would have been to delete these items; however, this would have compromised the clinical applicability of the scale. Another possibility would be to split the items and get different calibrations for men and women (59). This would make the scale more difficult to score, which, nonetheless, is justifiable if DIF results are replicated in further studies. DIF was no longer present in the top-down purification analysis, meaning that if DIF favors men for one item, to balance women are favored for another item. So, for the moment, and taking into consideration

that differences were of small magnitude, a conservative attitude was followed by avoiding scale modifications due to DIF. DIF did probably not influence the sex differences found in this study, since DIF refers to group differences at the same construct level.

The high floor effect was already discussed in the part of the discussion about the CTT. In Study 3, in which the total study sample was used, the floor effect was even more pronounced, as can be observed when comparing the distribution figures of both samples. As a consequence, a very asymmetrical person-item threshold distribution was found. Test performance would probably improve in a hospital-based sample, with a higher proportion of multimorbid patients and therefore less floor effects and better scale targeting.

RUMM2030 showed low reliability and although the PSI value improved when using RUMM2020, it still did not fit the criterion. This result is quite surprising since Cronbach´s alpha in the first study did fit the criterion. PSI and Cronbach´s alpha are interpreted in the same way and measure the extent to which the items measure the same construct. However, when distributions are skewed, the PSI gives a more accurate indication of internal consistency reliability (130). In the DBMA, disease burden is rated for single diseases. This makes items in the DBMA less related to each other than in 'regular' psychometric scales. However, the items do measure the same construct, i.e. disease burden. PSI in RUMM2020 is less influenced by floor effects than in RUMM2030, and was probably therefore closer to the criterion of >0.70. This could also make us expect the PSI to be higher in a sample with more multimorbidity. Low PSIs were also found in other widely used scales, such as the EQ-5D-3D (131).

According to the PCA, the DBMA was unidimensional. This result does not coincide with the exploratory factor analysis, which indentified five factors. In line with the literature, in

this case Rasch analysis provides superior results (132). The factor analysis was an exploratory analysis and was performed with the raw summative scores. Since these observations are non-linear, they can generate illusory factors (133). Moreover, in factor analysis items clustering at different performance levels are usually reported as different factors. Therefore, from factor analysis alone there is no way of knowing whether each factor is a dimension or part of a shared dimension. Although it makes sense to divide items into disease groups, the objective of the DBMA is to quantify disease burden caused by a series of different conditions on the whole and not per disease group.

The Rasch analysis provided a linear measure. Moderate correlations between the linear measure and physical functioning, perceived health and depression were found and, as well as a weak correlation with QoL. These correlations are lower than the correlations found in Study 1 and 2, which can be ascribed to the fact that a different sample was used in Study 3, with younger participants and an even higher health status. The relative precision analysis displayed some gain in precision in discriminating between age groups but when discriminating between men and women, no difference was found with the original scale. These data suggest that the linear measure is at least as valid as the raw summative score concerning discriminant validity.

The linear measure showed a peculiar distribution, due to the distances between lower DBMA scores: a raw score of 1 was converted into a linear score of 4.46; a raw score of 2 was converted into a linear value of 7.44. Moreover, RUMM2030 calculated parameter estimations for observations with missing values that did not coincide with the scores in the conversion table. The large distances between the lower scores can be interpreted in the following way: the difference in disease burden between raw scores of 0 and 1 is much bigger than the difference in disease burden between raw scores of 9 and 10. This can be illustrated with the classical example of studying for an exam: when using a 1-10 grading

system, it does not require the same amount of study time to pass from 0 to 1 (or 9 to 10) as to pass from 4 to 5. The same is the case for the response options of the DBMA.

Does this mean that from now on, when using the DBMA, the linear measure should be calculated? This depends on the context in which the DBMA is used. One of the main advantages of the DBMA is its simplicity. It repeats the same question for only the present conditions in a person which makes that it is very little time-consuming. The same is the case for the interpretation of the DBMA, which consists in simply summing the scores given to the individual conditions. Having to convert the summative score into a linear measure would make this process more complicated. Conversely, one of the advantages of a linear measure is that it allows the use of parametric statistics, given a normal distribution. In case of the DBMA, distributions will usually be skewed, even in study populations with high multimorbidity levels (45). Another advantage is that is allows to calculate change scores. The reason why this cannot be done with the raw score is because the measurement intervals between the response options are unequal (134), as explained above. Since these distances are unequal, the change in disease burden over time cannot be quantified because one point change does not always mean the same increase in disease burden along the construct continuum. Taking into account these advantages and disadvantages, it can be concluded that when calculating change scores, the linear measure should be applied. In other contexts, the raw summative score might be a good option.

In conclusion, an adequate fit to the Rasch model was achieved after rescoring the response options. Rasch analysis found the DBMA to be unidimensional and neither response dependency nor relevant DIF were found. The absence of DIF means that individuals with the same level of disease burden do not have different probabilities of attaining a certain score on the item. Scale targeting was below standards, with an

asymmetrical patient distribution and reliability was low. Further research in a hospital-based sample is proposed, in which the last two parameters are expected to improve. Rasch analysis provided a linear measure that permits the calculation of change scores. This makes the linear measure useful for clinical follow-up of patients as well as for comparisons before and after interventions to improve QoL in multimorbid patients.

5.4. LIMITATIONS AND STRENGTHS

5.4.1. LIMITATIONS

Some limitations should be acknowledged. Analyses in the first study were not corrected for complex samples or weights, since this would impede some of the feasibility and acceptability analyses. Rasch analysis does not permit correcting for complex samples or weights either, so in the third study, the uncorrected data were also used. Thus, results in these studies cannot be extrapolated to the Spanish population as a whole and cannot be compared with data in the second study.

Furthermore, as in other studies (45,47,48), the list of conditions included in the DBMA was adapted in the ELES-PS, which hinders comparisons across studies. Bayliss et al. (45) found a mean disease burden score of 20 in one of their studies, a value much higher that the mean score in the ELES-PS sample. However, since different lists of chronic conditions were used, these mean values are not comparable. A standard list of conditions should be developed for future use.

A third limitation was that it was not possible to study concurrent validity (the relation of the scale with other tests) since no other measures of multimorbidity or disease burden were included in the ELES-PS. Other authors compared the performance of the DBMA with the CCI, RxRisk score, Quan comorbidity index, CDS and CIRS (41,45,47), all with satisfactory results. It would have been of special interest to compare the DBMA as a predictor of mortality with another comorbidity index that was specifically validated to predict mortality, such as the CCI or the ECM.

Fourthly, as discussed before, a very healthy sample was used to validate a disease burden assessment. The ELES-PS sample showed a higher health status than the general Spanish population, as shown in Study 2. As a result, very high floor effects and a skewed distribution were found in Study 1. Study 2 showed a very low mortality rate and in Study 3 a very low PSI and an asymmetrical person-item threshold distribution were found. In a hospital-based sample, with higher multimorbidity and burden, these parameters would probably improve.

Fifthly and finally, the four-year follow-up time for mortality was quite short, especially given the high health status of the participants in the ELES-PS. Only 5.5% of the participants in the sample used in Study 2 had died, which led to limited statistical power in the predictive validity analyses. The same database might be used to repeat the analysed in future research, after a longer follow-up time.

5.4.2. STRENGHTS

This work used the data of a national survey, with a sample designed to be representative of the elderly population in Spain. Since the Study 2analyses were corrected for complex samples its generalizability to the general population is higher. The ELES-PS included a great variety of measures and indices, making it possible to assess the DBMA convergent validity with different self-reported outcomes. Moreover, it was possible to study the association with mortality, whose data were manually crossed with those of the National Death Index. Finally, this work consists of a quite complete psychometric validation, including a Rasch perspective. Rasch analysis cannot be performed in regular statistic programs such as Stata or SPSS and requires specific software and statistical expertise. It

evaluates psychometric attributes, such as DIF and category structure, that are not included in the CTT approach and provides a linear measure, with important advantages.

5.5. STUDY IMPLICATIONS

The DBMA is a scale that measures the subjective disease severity of 21 common chronic conditions. It repeats the same question for only present conditions, and due to this simple design, as well as being easy to understand and short administration time, it is especially useful for use in the older population.

Multimorbidity is an important public health problem and an essential aspect of care for chronically ill persons is to maximize their QoL. Different methods of measuring multimorbidity are associated with different health outcomes, and the choice of instrument depends on the outcome of interest. The DBMA has shown to be especially related to patient-centered outcomes as QoL and physical functioning, although it also showed significant relations with healthcare utilization and mortality.

The DBMA might be used in individual patients to assess and follow-up disease burden by chronic conditions. The scale may also be applied to identify patient groups with high disease burden for posterior interventions as well as the evaluation of these interventions. Some examples are interventions to improve QoL and functional status in multimorbid patients, such as self-management interventions, home care programs, or group meetings for chronically ill persons (135–137). The linear measure, as a result of the Rasch analysis in this work, may be applied for the calculation of change scores in these evaluations.

This work also identified areas for further research. Different psychometric parameters were expected to improve in hospital-based samples with more multimorbidity, such as floor effects, ITCC, scale targeting and reliability (PSI). Furthermore, the response scale structure should be confirmed in further research, and it would be interesting to also use a multimorbid population for this purpose. As stated in the limitations section, until now,

different authors have used different lists of conditions in the DBMA. A standard list of conditions should be developed in order to make results comparable across studies. Finally, to be able to choose the most suitable multimorbidity assessment instrument depending on the outcome of interest, a comparison of the DBMA with other multimorbidity measures in predicting mortality is suggested, which was not possible in the current research. Again, it would be interesting to do this in a less healthy sample, to find higher mortality rates, for which a longer follow-up time would be another valid option.

6. CONCLUSIONS

1. The DBMA consists of a list of chronic medical conditions from which respondents select the conditions they have. Limitations on daily life are rated on a scale from 1 to 5 for present conditions and non-present conditions are scored 0. Due to this design, the DBMA shows high floor effects per item, even in comorbid populations, since they reflect all absent conditions.

2. Scoring high for one condition does not mean the other conditions should be present as well. Therefore, items are not highly related to each other nor to the total score (low item homogeneity index and item-total correlations). There were no items linked in such way that the response to one item would determine the response to another (response dependency).

3. The DBMA was designed and validated to be associated with patient-centered outcomes. This was confirmed in this work: it showed satisfactory convergent validity with physical functioning, perceived health, depression, QoL and affect balance.

4. A positive association between the DBMA and healthcare use was found. The performance of the DBMA as a predictor of mortality had not been studied before and a positive association was found, despite the low mortality rate in the used sample.

5. Women scored higher on the DBMA not only because they had more present conditions, but also because they reported higher disease burden per condition. There were also differences by age, though less pronounced.

6. Exploratory factor analysis suggested the presence of five dimensions according to disease groups within the DBMA. However, Rasch analysis confirmed the scale to be unidimensional. The DBMA measures one dimension, namely the impact of chronic conditions on daily life, conceptualized as disease burden.

7. Reliability was satisfactory in CTT (Cronbach's alpha) but below standards in Rasch analysis (PSI). The high health status in the used sample might have played a role in this and further research is needed.

8. The participants were not able to make fine distinctions between response options (category structure), and rescoring of response options is proposed by Rasch analysis.

9. No relevant DIF was found for age, sex and educational level, meaning that persons belonging to different groups, with the same level of disease burden, did not rate their disease burden differently.

10. Rasch analysis provided a linear measure that should be used when calculating change scores.

In summary, results suggest that, despite some limitations such as reliability below the expected and high floor effects, the DBMA is an adequate patient-reported health outcome for measuring disease burden caused by 21 common chronic diseases in older adults. Persons with multiple chronic medical conditions are progressively becoming more common in our healthcare systems, and it is important to assess the impact that patients themselves experience because of their multimorbidity. The DBMA measures the burden of multimorbidity, by simply asking the patient to rate the impact of diseases on what is

most important to the patients themselves: everyday life. It is an easy-to-use tool that can contribute to the implementation of strategies to improve QoL and functional status for the growing group of older patients with multiple chronic conditions.

REFERENCES

1. United Nations, Department of Economic and Social Affairs, Population Division. World Population Ageing 2013 [Internet]. 2013 [cited 2015 Jun 26]. Available from: http://www.un.org/en/development/desa/population/publications/pdf/ageing/WorldPopulationAgeing2013.pdf

2. Rechel B, Grundy E, Robine J-M, Cylus J, Mackenbach JP, Knai C, et al. Ageing in the European Union. Lancet. 2013 Apr 13;381(9874):1312–22.

3. Newman AB. Comorbidity and Multimorbidity. In: Newman AB, Cauley JA, editors. The Epidemiology of Aging. Springer Netherlands; 2012. p. 119–33.

4. Marengoni A, Angleman S, Melis R, Mangialasche F, Karp A, Garmen A, et al. Aging with multimorbidity: A systematic review of the literature. Ageing Research Reviews. 2011 Sep;10(4):430–9.

5. Lenzi J, Avaldi VM, Rucci P, Pieri G, Fantini MP. Burden of multimorbidity in relation to age, gender and immigrant status: a cross-sectional study based on administrative data. BMJ Open. 2016 Dec 1;6(12):e012812.

6. Piccirillo JF, Vlahiotis A, Barrett LB, Flood KL, Spitznagel EL, Steyerberg EW. The Changing Prevalence of Comorbidity Across the Age Spectrum. Crit Rev Oncol Hematol. 2008 Aug;67(2):124–32.

7. Rizza A, Kaplan V, Senn O, Rosemann T, Bhend H, Tandjung R. Age- and gender-related prevalence of multimorbidity in primary care: the swiss fire project. BMC Fam Pract. 2012 Nov 24;13:113.

8. Fortin M, Stewart M, Poitras M-E, Almirall J, Maddocks H. A systematic review of prevalence studies on multimorbidity: toward a more uniform methodology. Ann Fam Med. 2012 Apr;10(2):142–51.

9. Sangha O, Stucki G, Liang MH, Fossel AH, Katz JN. The Self-Administered Comorbidity Questionnaire: a new method to assess comorbidity for clinical and health services research. Arthritis Rheum. 2003 Apr 15;49(2):156–63.

10. Forjaz MJ, Rodriguez-Blazquez C, Ayala A, Rodriguez-Rodriguez V, de Pedro-Cuesta J, Garcia-Gutierrez S, et al. Chronic conditions, disability, and quality of life in older adults with multimorbidity in Spain. Eur J Intern Med. 2015 Apr;26(3):176–81.

11. Nunes BP, Flores TR, Mielke GI, Thumé E, Facchini LA. Multimorbidity and mortality in older adults: A systematic review and meta-analysis. Arch Gerontol Geriatr. 2016 Dec;67:130–8.

12. Kim W, Song KY, Lee H-J, Han SU, Hyung WJ, Cho GS. The impact of comorbidity on surgical outcomes in laparoscopy-assisted distal gastrectomy: a retrospective analysis of multicenter results. Ann Surg. 2008 Nov;248(5):793–9.

13. Janssen-Heijnen MLG, Maas HAAM, Houterman S, Lemmens VEPP, Rutten HJT, Coebergh JWW. Comorbidity in older surgical cancer patients: Influence on patient care and outcome. European Journal of Cancer. 2007 Oct;43(15):2179–93.

14. Hewitt J, McCormack C, Tay HS, Greig M, Law J, Tay A, et al. Prevalence of multimorbidity and its association with outcomes in older emergency general surgical patients: an observational study. BMJ Open. 2016 Jan 3;6(3):e010126.

15. Culakova E, Poniewierski MS, Crawford J, Dale DC, Lyman GH. Impact of Comorbidities on Length of Stay and Mortality in Hospitalized Patients with Cancer and Febrile Neutropenia. Blood. 2014 Dec 6;124(21):2601–2601.

16. Glynn LG, Valderas JM, Healy P, Burke E, Newell J, Gillespie P, et al. The prevalence of multimorbidity in primary care and its effect on health care utilization and cost. Family Practice. 2011 Jan 10;28(5):516–23.

17. Bähler C, Huber CA, Brüngger B, Reich O. Multimorbidity, health care utilization and costs in an elderly community-dwelling population: a claims data based observational study. BMC Health Services Research. 2015;15:23.

18. Garin N, Olaya B, Moneta MV, Miret M, Lobo A, Ayuso-Mateos JL, et al. Impact of multimorbidity on disability and quality of life in the Spanish older population. PLoS ONE. 2014;9(11):e111498.

19. Marventano S, Ayala A, Gonzalez N, Rodríguez-Blázquez C, Garcia-Gutierrez S, Forjaz MJ, et al. Multimorbidity and functional status in community-dwelling older adults. Eur J Intern Med. 2014 Sep;25(7):610–6.

20. Marventano S, Ayala A, Gonzalez N, Rodríguez-Blázquez C, Garcia-Gutierrez S, Forjaz MJ. Multimorbidity and functional status in institutionalized older adults. European Geriatric Medicine. 2016 Feb 1;7(1):34–9.

21. St John PD, Tyas SL, Menec V, Tate R. Multimorbidity, disability, and mortality in community-dwelling older adults. Can Fam Physician. 2014 May;60(5):e272-280.

22. Fortin M, Lapointe L, Hudon C, Vanasse A, Ntetu AL, Maltais D. Multimorbidity and quality of life in primary care: a systematic review. Health Qual Life Outcomes. 2004 Sep 20;2:51.

23. Yurkovich M, Avina-Zubieta JA, Thomas J, Gorenchtein M, Lacaille D. A systematic review identifies valid comorbidity indices derived from administrative health data. J Clin Epidemiol. 2015 Jan;68(1):3–14.

24. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis. 1987;40(5):373–83.

25. Quan H, Li B, Couris CM, Fushimi K, Graham P, Hider P, et al. Updating and Validating the Charlson Comorbidity Index and Score for Risk Adjustment in Hospital Discharge Abstracts Using Data From 6 Countries. Am J Epidemiol. 2011 Mar 15;173(6):676–82.

26. Charlson ME, Charlson RE, Peterson JC, Marinopoulos SS, Briggs WM, Hollenberg JP. The Charlson comorbidity index is adapted to predict costs of chronic disease in primary care patients. J Clin Epidemiol. 2008 Dec;61(12):1234–40.

27. Kieszak SM, Flanders WD, Kosinski AS, Shipp CC, Karp H. A comparison of the Charlson comorbidity index derived from medical record data and administrative billing data. J Clin Epidemiol. 1999 Feb;52(2):137–42.

28. Habbous S, Chu KP, Harland LTG, La Delfa A, Fadhel E, Sun B, et al. Validation of a one-page patient-reported Charlson comorbidity index questionnaire for upper aerodigestive tract cancer patients. Oral Oncol. 2013 May;49(5):407–12.

29. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. Med Care. 1998 Jan;36(1):8–27.

30. Moltó A, Dougados M. Comorbidity indices. Clin Exp Rheumatol. 2014 Oct;32(5 Suppl 85):S-131-134.

31. Gutacker N, Bloor K, Cookson R. Comparing the performance of the Charlson/Deyo and Elixhauser comorbidity measures across five European countries and three conditions. Eur J Public Health. 2015 Feb;25 Suppl 1:15–20.

32. Lieffers JR, Baracos VE, Winget M, Fassbender K. A comparison of Charlson and Elixhauser comorbidity measures to predict colorectal cancer survival using administrative health data. Cancer. 2011 May 1;117(9):1957–65.

33. Menendez ME, Neuhaus V, van Dijk CN, Ring D. The Elixhauser comorbidity method outperforms the Charlson index in predicting inpatient death after orthopaedic surgery. Clin Orthop Relat Res. 2014 Sep;472(9):2878–86.

34. Linn BS, Linn MW, Gurel L. Cumulative illness rating scale. J Am Geriatr Soc. 1968 May;16(5):622–6.

35. Miller MD, Paradis CF, Houck PR, Mazumdar S, Stack JA, Rifai AH, et al. Rating chronic medical illness burden in geropsychiatric practice and research: application of the Cumulative Illness Rating Scale. Psychiatry Res. 1992 Mar;41(3):237–48.

36. de Groot V, Beckerman H, Lankhorst GJ, Bouter LM. How to measure comorbidity. a critical review of available methods. J Clin Epidemiol. 2003 Mar;56(3):221–9.

37. Olomu AB, Corser WD, Stommel M, Xie Y, Holmes-Rovner M. Do self-report and medical record comorbidity data predict longitudinal functional capacity and quality of life health outcomes similarly? BMC Health Services Research. 2012 Nov 14;12(1):398.

38. Motl RW, Fernhall B, McAuley E, Cutter G. Physical activity and self-reported cardiovascular comorbidities in persons with multiple sclerosis: evidence from a cross-sectional analysis. Neuroepidemiology. 2011;36(3):183–91.

39. Fiest KM, Currie SR, Williams JVA, Wang J. Chronic conditions and major depression in community-dwelling older adults. J Affect Disord. 2011 Jun;131(1–3):172–8.

40. Robinski M, Strich F, Mau W, Girndt M. Validating a Patient-Reported Comorbidity Measure with Respect to Quality of Life in End-Stage Renal Disease. PLOS ONE. 2016 Jun 13;11(6):e0157506.

41. Bayliss EA, Ellis JL, Steiner JF. Subjective assessments of comorbidity correlate with quality of life health outcomes: initial validation of a comorbidity assessment instrument. Health Qual Life Outcomes. 2005;3:51.

42. Scarborough P, Bhatnagar P, Wickramasinghe KK, Allender S, Foster C, Rayner M. The economic burden of ill health due to diet, physical inactivity, smoking, alcohol and

obesity in the UK: an update to 2006-07 NHS costs. J Public Health (Oxf). 2011 Dec;33(4):527–35.

43. Pinheiro P, Plaß D, Krämer A. The Burden of Disease Approach for Measuring Population Health. In: Krämer A, Khan MH, Kraas F, editors. Health in Megacities and Urban Areas. Heidelberg: Physica-Verlag; 2011. p. 21–38.

44. Bayliss EA, Ellis JL, Steiner JF. Barriers to Self-Management and Quality-of-Life Outcomes in Seniors With Multimorbidities. Ann Fam Med. 2007 Sep;5(5):395–402.

45. Bayliss EA, Ellis JL, Steiner JF. Seniors' self-reported multimorbidity captured biopsychosocial factors not incorporated in two other data-based morbidity measures. J Clin Epidemiol. 2009 May;62(5):550–7.e1.

46. Bayliss EA, Ellis JL, Shoup JA, Zeng C, McQuillan DB, Steiner JF. Association of patient-centered outcomes with patient-reported and ICD-9-based morbidity measures. Ann Fam Med. 2012 Apr;10(2):126–33.

47. Poitras M-E, Fortin M, Hudon C, Haggerty J, Almirall J. Validation of the disease burden morbidity assessment by self-report in a French-speaking population. BMC Health Services Research. 2012;12(1):35.

48. Hudon C, Fortin M, Poitras M-E, Almirall J. The relationship between literacy and multimorbidity in a primary care setting. BMC Family Practice. 2012 Apr 26;13(1):33.

49. Robichaud-Hallé L, Beaudry M, Fortin M. Obstructive sleep apnea and multimorbidity. BMC Pulmonary Medicine. 2012;12:60.

50. Rodriguez-Blazquez C. Sistematización de la metodología de validación e interpretación de resultados para medidas clínicas de la enfermedad de Parkinson. [Madrid]: Universidad Autónoma de Madrid; 2014.

51. Cappelleri JC, Lundy JJ, Hays RD. Overview of Classical Test Theory and Item Response Theory for Quantitative Assessment of Items in Developing Patient-Reported Outcome Measures. Clin Ther. 2014 May;36(5):648–62.

52. Petrillo J, Cano SJ, McLeod LD, Coon CD. Using Classical Test Theory, Item Response Theory, and Rasch Measurement Theory to Evaluate Patient-Reported Outcome Measures: A Comparison of Worked Examples. Value in Health. 2015 Jan 1;18(1):25–34.

53. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010 Jul;63(7):737–45.

54. Thurstone LL. A law of comparative judgment. Psychological Review. 1927;34(4):273–86.

55. Rasch G. On General Laws and the Meaning of Measurement in Psychology. Danmarks pædagogiske Institut; 1961. 13 p.

56. Forjaz MJ, Prieto-Flores M-E, Ayala A, Rodriguez-Blazquez C, Fernandez-Mayoralas G, Rojo-Perez F, et al. Measurement properties of the Community Wellbeing Index in older adults. Qual Life Res. 2011 Jun;20(5):733–43.

57. Tennant A, McKenna SP, Hagell P. Application of Rasch Analysis in the Development and Application of Quality of Life Instruments. Value in Health. 2004 Sep 1;7:S22–6.

58. Forjaz MJ, Ayala A, Rodriguez-Blazquez C, Prieto-Flores M-E, Fernandez-Mayoralas G, Rojo-Perez F, et al. Rasch analysis of the International Wellbeing Index in older adults. Int Psychogeriatr. 2012 Feb;24(2):324–32.

59. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: What is it and why use it? When should it be applied, and what should one look for in a Rasch paper? Arthritis & Rheumatism. 2007 Dec 15;57(8):1358–62.

60. Wijers IGM, Ayala A, Rodriguez-Blazquez C, Rodriguez-Laso A, Rodriguez-Rodriguez V, Forjaz MJ. Disease burden morbidity assessment by self-report: Psychometric properties in older adults in Spain. Geriatr Gerontol Int. 2016 Jul 18;

61. Wijers IGM, Ayala A, Rodriguez-Blazquez C, Rodriguez-Laso A, Rodriguez-Rodriguez V, Forjaz MJ. Rasch analysis and construct validity of the Disease Burden Morbidity Assessment in older adults. The Gerontologist. 2017;In press.

62. Teófilo Rodríguez J, González Cabezas AN, Díaz Veiga P, Rodríguez Rodríguez V. Estudio Longitudinal Envejecer en España: El proyecto ELES. Perfiles y tendencias. 2011;50:1–44.

63. Linacre JM. Sample Size and Item Calibration Stability. Rasch Measurement Transactions. 1994;7(4):328.

64. Linacre JM. A User´s Guide to Winsteps Ministep Rasch-Model Computer Programs [Internet]. 2016 [cited 2017 Feb 5]. Available from: http://www.winsteps.com/manuals.htm

65. Mavranezouli I, Brazier JE, Young TA, Barkham M. Using Rasch analysis to form plausible health states amenable to valuation: the development of CORE-6D from a measure of common mental health problems (CORE-OM). Qual Life Res. 2011 Apr;20(3):321–33.

66. Smith AB, Rush R, Fallowfield LJ, Velikova G, Sharpe M. Rasch fit statistics and sample size considerations for polytomous data. BMC Med Res Methodol. 2008 May 29;8:33.

67. Byles JE, D'Este C, Parkinson L, O'Connell R, Treloar C. Single index of multimorbidity did not predict multiple outcomes. J Clin Epidemiol. 2005 Oct;58(10):997–1005.

68. Fried LP, Bandeen-Roche K, Kasper JD, Guralnik JM. Association of comorbidity with disability in older women: the Women's Health and Aging Study. J Clin Epidemiol. 1999 Jan;52(1):27–37.

69. Groll DL, To T, Bombardier C, Wright JG. The development of a comorbidity index with physical function as the outcome. J Clin Epidemiol. 2005 Jun;58(6):595–602.

70. Ramond-Roquin A, Haggerty J, Lambert M, Almirall J, Fortin M. Different Multimorbidity Measures Result in Varying Estimated Levels of Physical Quality of Life in Individuals with Multimorbidity: A Cross-Sectional Study in the General Population. BioMed Research International. 2016 Mar 16;2016:e7845438.

71. Irwin M, Artin K, Oxman MN. Screening for depression in the older adult: Criterion validity of the 10-item center for epidemiological studies depression scale (ces-d). Arch Intern Med. 1999 Aug 9;159(15):1701–4.

72. Robison J, Gruman C, Gaztambide S, Blank K. Screening for depression in middle-aged and older puerto rican primary care patients. J Gerontol A Biol Sci Med Sci. 2002 May;57(5):M308-314.

73. The International Wellbeing Group. Personal Wellbeing Index-Adult (PWI-A). Manual 2013. [Internet]. 5th ed. Melbourne: Australian Centre on Quality of Life, Deakin University; 2013 [cited 2013 Dec 11]. 41p p. Available from: http://www.deakin.edu.au/research/acqol/instruments/wellbeing-index/pwi-a-english.pdf

74. Fernandez-Martinez B, Prieto-Flores M-E, Forjaz MJ, Fernández-Mayoralas G, Rojo-Pérez F, Martínez-Martín P. Self-perceived health status in older adults: regional and sociodemographic inequalities in Spain. Revista de Saúde Pública. 2012 Apr;46(2):310–9.

75. Health and Retirement Study [Internet]. [cited 2014 Oct 27]. Available from: http://hrsonline.isr.umich.edu/index.php

76. Diener E, Wirtz D, Tov W, Kim-Prieto C, Choi D, Oishi S, et al. New Well-being Measures: Short Scales to Assess Flourishing and Positive and Negative Feelings. Soc Indic Res. 2009 May 28;97(2):143–56.

77. Martinez-Martin P, Forjaz MJ. How to Evaluate Validation Data. In: Sampaio C, Goetz CG, Schrag A, editors. Rating Scales in Parkinson's Disease. Oxford University Press; 2012. p. 16–41.

78. Virués-Ortega J, Rodríguez-Blázquez C, Micheli F, Carod-Artal FJ, Serrano-Dueñas M, Martínez-Martín P. Cross-cultural evaluation of the modified Parkinson Psychosis Rating Scale across disease stages. Mov Disord. 2010 Jul 30;25(10):1391–8.

79. Diaz-Redondo A, Rodriguez-Blazquez C, Ayala A, Martinez-Martin P, Forjaz MJ, Spanish Research Group on Quality of Life and Aging. EQ-5D rated by proxy in institutionalized older adults with dementia: psychometric pros and cons. Geriatr Gerontol Int. 2014 Apr;14(2):346–53.

80. Gijsen R, Hoeymans N, Schellevis FG, Ruwaard D, Satariano WA, van den Bos GA. Causes and consequences of comorbidity: a review. J Clin Epidemiol. 2001 Jul;54(7):661–74.

81. Caughey GE, Vitry AI, Gilbert AL, Roughead EE. Prevalence of comorbidity of chronic diseases in Australia. BMC Public Health. 2008 Jun 27;8(1):221.

82. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2 edition. Hillsdale, N.J: Routledge; 1988.

83. Diedenhofen B, Musch J. cocor: A Comprehensive Solution for the Statistical Comparison of Correlations. PLoS ONE. 2015 Apr 2;10(4):e0121945.

84. Violan C, Foguet-Boreu Q, Flores-Mateo G, Salisbury C, Blom J, Freitag M, et al. Prevalence, Determinants and Patterns of Multimorbidity in Primary Care: A Systematic Review of Observational Studies. PLoS ONE. 2014 Jul 21;9(7):e102149.

85. Henry C, M'Bailara K, Mathieu F, Poinsot R, Falissard B. Construction and validation of a dimensional scale exploring mood disorders: MAThyS (Multidimensional Assessment of Thymic States). BMC Psychiatry. 2008;8:82.

86. Newson R. Parameters behind "nonparametric" statistics: Kendall's tau, Somers' D and median differences. The Stata Journal. 2(1):45–64.

87. Cummins RA. Subjective Wellbeing, Homeostatically Protected Mood and Depression: A Synthesis. J Happiness Stud. 2009 Oct 9;11(1):1–17.

88. Bursac Z, Gauss CH, Williams DK, Hosmer DW. Purposeful selection of variables in logistic regression. Source Code Biol Med. 2008 Dec 16;3:17.

89. Vanagas G. Receiver operating characteristic curves and comparison of cardiac surgery risk stratification systems. Interact Cardiovasc Thorac Surg. 2004 Jun;3(2):319–22.

90. RUMM Laboratory. RUMM 2030 [Internet]. 2014 [cited 2017 Apr 14]. Available from: http://www.rummlab.com.au/

91. Masters GN. A Rasch Model for Partial Credit Scoring. Psychometrika. 1982;47(2):149–74.

92. Pallant JF, Tennant A. An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). Br J Clin Psychol. 2007 Mar;46(Pt 1):1–18.

93. Cronbach's α and the Person Separation Index (PSI) [Internet]. [cited 2015 Oct 23]. Available from: http://www.rummlab.com.au/rmrelidx2030.pdf

94. Smith AB, Wright EP, Rush R, Stark DP, Velikova G, Selby PJ. Rasch analysis of the dimensional structure of the Hospital Anxiety and Depression Scale. Psychooncology. 2006 Sep;15(9):817–27.

95. Tennant A, Pallant JF. DIF matters: A practical approach to test if Differential Item Functioning makes a difference. Rasch Measurement Transactions. 2007;20(4):1082–4.

96. Tennant A, Penta M, Tesio L, Grimby G, Thonnard J-L, Slade A, et al. Assessing and adjusting for cross-cultural validity of impairment and activity limitation scales through differential item functioning within the framework of the Rasch model: the PRO-ESOR project. Med Care. 2004 Jan;42(1 Suppl):I37-48.

97. Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. Lancet. 2012 Jul 7;380(9836):37–43.

98. Global Burden of Disease Study 2013 Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet. 2015 Aug 22;386(9995):743–800.

99. Las Hayas C, Bilbao A, Quintana JM, Garcia S, Lafuente I, IRYSS Cataract Group. A comparison of standard scoring versus Rasch scoring of the visual function index-14 in patients with cataracts. Invest Ophthalmol Vis Sci. 2011 Jun;52(7):4800–7.

100. Sakthong P, Charoenvisuthiwongs R, Shabunthom R. A comparison of EQ-5D index scores using the UK, US, and Japan preference weights in a Thai sample with type 2 diabetes. Health Qual Life Outcomes. 2008;6:71.

101. Fitzpatrick R, Norquist JM, Jenkinson C, Reeves BC, Morris RW, Murray DW, et al. A comparison of Rasch with Likert scoring to discriminate between patients' evaluations of total hip replacement surgery. Qual Life Res. 2004 Mar;13(2):331–8.

102. Instituto Nacional de Estadística. Population and Housing Census 2011 [Internet]. [cited 2016 Oct 26]. Available from: http://www.ine.es/en/censos2011_datos/cen11_datos_inicio_en.htm

103. National Health Survey 2011-2012 [Internet]. [cited 2016 Oct 26]. Available from: http://ine.es/dynt3/inebase/en/index.htm?type=pcaxis&path=/t15/p419/a2011/p01/&file=pcaxis

104. Instituto Nacional de Estadística. Mortality tables for the population of Spain 1991-2013 [Internet]. [cited 2016 Oct 26]. Available from: http://www.ine.es/jaxi/tabla.do?path=/t20/p319a/serie/p01/l1/&file=01001.px&type=pcaxis&L=1

105. Rodríguez Laso Á, Urdaneta Artola E, de la Fuente Sánchez M, Galindo Moreno E, Yanguas Lezáun JJ, Rodríguez Rodríguez V. Análisis del sesgo de selección en el piloto de un estudio longitudinal sobre envejecimiento en España. Gaceta Sanitaria. 2013 Sep;27(5):425–32.

106. Katz S, Calasanti T. Critical Perspectives on Successful Aging: Does It "Appeal More Than It Illuminates"? Gerontologist. 2015 Feb;55(1):26–33.

107. Kurella M, Ireland C, Hlatky MA, Shlipak MG, Yaffe K, Hulley SB, et al. Physical and sexual function in women with chronic kidney disease. Am J Kidney Dis. 2004 May;43(5):868–76.

108. Fried LF, Lee JS, Shlipak M, Chertow GM, Green C, Ding J, et al. Chronic kidney disease and functional limitation in older people: health, aging and body composition study. J Am Geriatr Soc. 2006 May;54(5):750–6.

109. Bowling CB, Sawyer P, Campbell RC, Ahmed A, Allman RM. Impact of Chronic Kidney Disease on Activities of Daily Living in Community-Dwelling Older Adults. J Gerontol A Biol Sci Med Sci. 2011 Jun;66A(6):689–94.

110. Fraser DD, Kee CC, Minick P. Living with chronic obstructive pulmonary disease: insiders' perspectives. J Adv Nurs. 2006 Sep;55(5):550–8.

111. Williams V, Bruton A, Ellis-Hill C, McPherson K. What really matters to patients living with chronic obstructive pulmonary disease? An exploratory study. Chron Respir Dis. 2007;4(2):77–85.

112. Bourbeau J. Activities of life: the COPD patient. COPD. 2009 Jun;6(3):192–200.

113. Moritz DJ, Kasl SV, Berkman LF. Cognitive Functioning and the Incidence of Limitations in Activities of Daily Living in an Elderly Community Sample. Am J Epidemiol. 1995 Jan 1;141(1):41–9.

114. Gure TR, Kabeto MU, Plassman BL, Piette JD, Langa KM. Differences in Functional Impairment Across Subtypes of Dementia. J Gerontol A Biol Sci Med Sci. 2010 Apr;65A(4):434–41.

115. Gadermann AM, Alonso J, Vilagut G, Zaslavsky AM, Kessler RC. Comorbidity and disease burden in the National Comorbidity Survey Replication (NCS-R). Depress Anxiety. 2012 Sep;29(9):797–806.

116. Moussavi S, Chatterji S, Verdes E, Tandon A, Patel V, Ustun B. Depression, chronic diseases, and decrements in health: results from the World Health Surveys. The Lancet. 2007 Sep 14;370(9590):851–8.

117. Wensing M, Vingerhoets E, Grol R. Functional status, health problems, age and comorbidity in primary care patients. Qual Life Res. 2001;10(2):141–8.

118. Cheng L, Cumber S, Dumas C, Winter R, Nguyen KM, Nieman LZ. Health related quality of life in pregeriatric patients with chronic diseases at urban, public supported clinics. Health Qual Life Outcomes. 2003;1:63.

119. Fortin M, Hudon C, Dubois M-F, Almirall J, Lapointe L, Soubhi H. Comparative assessment of three different indices of multimorbidity for studies on health-related quality of life. Health Qual Life Outcomes. 2005;3:74.

120. Jindai K. Multimorbidity and Functional Limitations Among Adults 65 or Older, NHANES 2005–2012. Prev Chronic Dis [Internet]. 2016 [cited 2017 Mar 4];13. Available from: https://www.cdc.gov/pcd/issues/2016/16_0174.htm

121. Prados-Torres A, Calderón-Larrañaga A, Hancco-Saavedra J, Poblador-Plou B, van den Akker M. Multimorbidity patterns: a systematic review. J Clin Epidemiol. 2014 Mar;67(3):254–66.

122. Hunt SM, McEwen J, McKenna SP. Perceived health: age and sex comparisons in a community. J Epidemiol Community Health. 1984 Jun;38(2):156–60.

123. Henchoz K, Cavalli S, Girardin M. Health perception and health status in advanced old age: A paradox of association. Journal of Aging Studies. 2008 Aug;22(3):282–90.

124. Pinquart M. Correlates of subjective health in older adults: a meta-analysis. Psychol Aging. 2001 Sep;16(3):414–26.

125. Windsor TD, Burns RA, Byles JE. Age, physical functioning, and affect in midlife and older adulthood. J Gerontol B Psychol Sci Soc Sci. 2013 May;68(3):395–9.

126. Mehta A, Malley B, Walkey A. Formulating the Research Question. In: Secondary Analysis of Electronic Health Records [Internet]. Springer International Publishing; 2016 [cited 2017 Mar 21]. p. 81–92. Available from: http://link.springer.com/book/10.1007/978-3-319-43742-2

127. Dalen MT van, Suijker JJ, MacNeil-Vroomen J, Rijn M van, Charante EPM van, Rooij SE de, et al. Self-Report of Healthcare Utilization among Community-Dwelling Older Persons: A Prospective Cohort Study. PLOS ONE. 2014 Apr 7;9(4):e93372.

128. Hobart JC, Cano SJ, Zajicek JP, Thompson AJ. Rating scales as outcome measures for clinical trials in neurology: problems, solutions, and recommendations. Lancet Neurol. 2007 Dec;6(12):1094–105.

129. Young T, Yang Y, Brazier JE, Tsuchiya A, Coyne K. The first stage of developing preference-based measures: constructing a health-state classification using Rasch analysis. Qual Life Res. 2009 Mar;18(2):253–65.

130. Kurtaiş Y, Őztuna D, Küçükdeveci AA, Kutlay Ş, Hafiz M, Tennant A. Reliability, construct validity and measurement potential of the ICF comprehensive core set for osteoarthritis. BMC Musculoskelet Disord. 2011 Nov 8;12:255.

131. Pickard AS, De Leon MC, Kohlmann T, Cella D, Rosenbloom S. Psychometric comparison of the standard EQ-5D to a 5 level version in cancer patients. Med Care. 2007 Mar;45(3):259–63.

132. Salzberger T. The Rasch model and factor analysis: Complementary or mutually exclusive? Rasch Measurement Transactions. 26(3):1373–8.

133. Linacre JM. Rasch Analysis first or Factor Analysis first? Rasch Measurement Transactions. 1998;11(4):603.

134. Kissane BV. The Measurement of Change as the Study of the Rate of Change. Education Research and Perspectives. 1982;9(1):55–72.

135. Elzen H, Slaets JPJ, Snijders TAB, Steverink N. Evaluation of the chronic disease self-management program (CDSMP) among chronically ill older people in the Netherlands. Soc Sci Med. 2007 May;64(9):1832–41.

136. Levine S, Steinman BA, Attaway K, Jung T, Enguidanos S. Home care program for patients at high risk of hospitalization. Am J Manag Care. 2012 Aug 1;18(8):e269-276.

137. Scott JC, Conner DA, Venohr I, Gade G, McKenzie M, Kramer AM, et al. Effectiveness of a group outpatient visit model for chronically ill older health maintenance organization members: a 2-year randomized trial of the cooperative health care clinic. J Am Geriatr Soc. 2004 Sep;52(9):1463–70.

APPENDICES

APPENDIX I: Article of Study 1

APPENDIX II: Manuscript of Study 3

Geriatrics Gerontology INTERNATIONAL Medline Indexed

Geriatr Gerontol Int 2016

ORIGINAL ARTICLE: EPIDEMIOLOGY, CLINICAL PRACTICE AND HEALTH

Disease burden morbidity assessment by self-report: Psychometric properties in older adults in Spain

Irene GM Wijers,[1] Alba Ayala,[2] Carmen Rodriguez-Blazquez,[3] Angel Rodriguez-Laso,[4] Vicente Rodriguez-Rodriguez[5] and Maria João Forjaz[2]

[1]Department of Preventive Medicine and Quality Management, General University Hospital Gregorio Marañón, [2]National School of Public Health, Institute of Health Carlos III and REDISSEC, [3]National Center of Epidemiology, Carlos III Institute of Health and CIBERNED, [4]Department of Pharmacology and Biotechnology, Universidad Europea, and [5]Institute of Economics, Geography and Demography, Spanish National Research Council, Madrid, Spain

Aim: To carry out an analysis of the psychometric properties of the Disease Burden Morbidity Assessment (DBMA) according to the assumptions of the Classical Test Theory.

Methods: A sample of 707 community-dwelling adults aged 65 years and older, living in Spain, completed the DBMA. Psychometric properties of the scale (feasibility, acceptability, scaling assumptions, reliability and construct validity) were analyzed.

Results: The mean DBMA score was 6.8. Feasibility and acceptability were satisfactory, except for large floor effects (>50%), as well as a skewed distribution (1.8). Item-total corrected correlation ranged 0.10–0.49, item homogeneity index was 0.09 and Cronbach's alpha was 0.72. Disease burden correlated strongly with physical functioning ($r=-0.56$) and perceived health ($r=-0.56$), and moderately with depression ($r=0.41$) and the Personal Wellbeing Index ($r=-0.41$). Exploratory factor analysis extracted five factors, explaining 44% of the variance.

Conclusions: The DBMA is an acceptable and valid instrument for measuring disease burden in older adults. Future studies should include Rasch analysis to further assess dimensionality and explore other measurement properties. **Geriatr Gerontol 2016; ••: ••–••**

Keywords: aged, burden of illness, chronic disease, comorbidity, psychometrics.

Introduction

As chronic health conditions are more frequent among older people, the process of population aging will further increase the prevalence of these, as well as the prevalence of co- and multimorbidity.[1] The terms co- and multimorbidity both refer to the presence of multiple chronic conditions in one person, but in the case of comorbidity, these are studied from the perspective of one index-disease.[2] The term, multimorbidity, is currently preferred when referring to measures of chronic health conditions, and will therefore be used in this study, although both terms could be used interchangeably.

Multimorbidity is a frequent phenomenon in elderly populations, with an estimated prevalence ranging from 55 to 98% in persons aged 65 years and older.[3] It is an important prognostic factor, with well-described negative effects on mortality, complications, surgical outcome and hospital length of stay,[4] and a direct and independent effect on disability and quality of life (QoL).[5]

The Charlson Comorbidity Index[6] and the Elixhauser's comorbidity tool[7] are two widely applied tools for multimorbidity risk adjustment. Both scales weigh disease categories based on the associated risk of mortality. However, these tools do not assess disease severity, and studies have shown that the incorporation of severity assessment significantly improves the predictive value of morbidity measures.[4,8]

Bayliss *et al.* developed a multimorbidity scale, the Disease Burden Morbidity Assessment (DBMA) that assesses the impact of 23 conditions on daily activities as a measure of disease severity, conceptualized as self-reported disease burden.[9] This scale was designed in order to create a subjective measure of comorbidity, to be used especially in studies using QoL outcomes, where the patient's perception plays an important role. The

Accepted for publication 26 April 2016.

Correspondence: Dr Maria Joao Forjaz PhD, Escuela Nacional de Sanidad, Avda. Monforte de Lemos 5, 28029 Madrid, Spain. Email: jforjaz@isciii.es

© 2016 Japan Geriatrics Society

developing authors carried out an initial validation of this scale in older adults, calculating sensitivities and specificities relative to chart review, and assessing convergent validity of the DBMA with QoL outcomes. In following studies, the association between the DBMA and biopsychosocial factors were analyzed.[10–12] Poitras *et al.* validated the DBMA in French, and assessed test–retest reliability and concurrent validity with the cumulative illness rating scale.[13]

A complete analysis of psychometric properties, following the steps of the Classical Test Theory, had not been carried out yet. The aim of the present investigation was to complement the validation process by studying these properties in a sample of older adults. A secondary goal was to assess the performance of this scale in Spanish. To the best of our knowledge, this is the first publication about this scale applied to a Spanish-speaking population, and adds information about the use of the DBMA in a different language and cultural setting.

Methods

Study design

Data were extracted from the Aging in Spain Longitudinal Study, Pilot Survey (ELES-PS). This study was carried out in 2011 among community-dwelling adults aged 50 years or older, living in Spain. Representative participants were randomly selected on a national geographic basis, in three stages. First, stratified clusters were formed by autonomous region and municipality size, proportionally to its population aged 50 years and older. Second stage units consisted of households with telephone lines that were selected from census data. Finally, a random selection of individuals was carried out from each household, with post-stratification by sex and age group (50–59, 60–69, 70–79 and 80–89 years). People from the Basque region were overrepresented in the sample.

Data were collected in four phases: a telephone questionnaire, a visit by trained nurses, a Computer-Assisted Personal Interviewing (CAPI) questionnaire and a self-administered questionnaire. The DBMA was included in the CAPI questionnaire, applied to 1400 persons. As this scale was originally developed to be used in older adults, we selected a subsample of persons aged 65 years and older, resulting in a final sample of 707 persons.

The CAPI interviews took place at the respondents' homes, and were carried out by a trained interviewer. In the case of cognitive impairment, the interview was carried out with a proxy respondent, excluding in this case the questions that required a subjective judgement. Questions about sensitive subjects, such as sex life and depression, were administered through the self-administered questionnaire.

Assessments

We applied a scale consisting of a list of 21 medical diagnoses of chronic conditions whose severity was scored following the procedure of the DBMA instrument described by Bayliss *et al.*[9] Participants were asked for every condition whether they had it, and if so, to what extent these conditions limited their everyday activities on a scale from 1 ("not at all") to 5 ("a lot"). Non-present conditions were scored 0. The DBMA total score, representing the level of self-reported disease burden, was a result of the sum of the limitation levels assigned to the 21 conditions.

As in other studies,[12–14] we adapted the original DBMA's 23-item list, based on the conditions used in other multimorbidity indices.[4,9,15–17] From the publications in which analysis per condition had been carried out, we selected those conditions that specifically predicted mortality, hospitalization or future handicaps and those that showed a transversal association with physical functioning.[15–17] In the case where analysis per condition had not been carried out, all conditions were selected.[4,9] As a criterion, only conditions selected from more than one index were included. A few exceptions were made: liver diseases were not included because of their low prevalence, and urinary tract conditions, anxiety and memory-related disorders were added because of their high prevalence in older adults.

To screen for depression, a dichotomous 10-item Center for Epidemiologic Studies Depression Scale was administered through the self-administered questionnaire.[18] This scale was validated for the Spanish-speaking population, and has shown to be a valid instrument for detecting major depression in older adults. It contains 10 questions with "yes/no" response categories, and a score of 1 is assigned for every positive answer for depression. A sum score of 3 was used as a cut-off point for depression.

A 24-item scale, as used in the Health and Retirement Study, was included in the CAPI questionnaire as a measure of physical functioning.[19] This scale consists of a list of 24 different activities (e.g. getting dressed, walking 100 m and making phone calls) and participants are asked whether they experience difficulties when carrying them out, on a scale from 1 (always) to 4 (never). Scores were summed in order to obtain a measure of physical functioning.

QoL was assessed through the Personal Wellbeing Index (PWI),[20] included in the CAPI-questionnaire, in which respondents are asked to grade, on a scale of 1 to 10, their satisfaction with seven life dimensions: standard of living, personal health, achieving in life, personal relationships, personal safety, community-connectedness and future security. Total subscores were lineally transformed into a 0–100 scale, and higher total scores indicate better QoL.[21] The second dimension of the PWI, grading personal health on a scale of 1 to 10, was used as a measure of perceived health.

© 2016 Japan Geriatrics Society

Data analysis

We examined the following psychometric properties: feasibility, acceptability, scaling assumptions, reliability and construct validity. Feasibility was assessed by determining the percentage of missing values per item and the percentage of computable scores for the total scale, considering acceptable scores <10% and >90%, respectively.[22] Acceptability was explored by comparing possible and observed scores, and assessing mean-to-median difference for the total scale (criterion, <10% of the scale range) as well as floor and ceiling effects (<15%) and skewness (–1 to 1).[23]

Scaling assumptions were determined through the item-total corrected correlation for each item (criterion $r \geq 0.40$).[24] Reliability was assessed through internal consistency (Cronbach's alpha) and the item homogeneity index (criteria: $\alpha \geq 0.70$ and $r \geq 0.30$, respectively).[22]

For convergent validity, we expected self-reported disease burden to be negatively associated with perceived health, physical functioning and QoL, and to find a positive association with depression.[1,5,25] This was calculated through Spearman's rank correlation coefficients because of the non-normal distribution. Correlation coefficients were interpreted following Cohen's conventions, considering correlations ≥0.5 large, 0.5–0.3 moderate and 0.3–0.1 small magnitudes of effect.[26] Spearman's rank correlations were repeated using the self-reported number of conditions instead of disease burden, to evaluate the added value of assessing the impact of conditions on daily life. These correlations were compared using a Fisher's r-to-z transformation.[27] Because the DBMA had a higher number of missing values than the disease count variable, we excluded those cases with missing values for DBMA in this analysis in order to make the two variables comparable. Known-groups validity was examined comparing disease burden by sex and age groups (Wilcoxon rank-sum test). As multimorbidity is more frequent in women and older people, we hypothesized to find significantly higher scores in these groups.[28]

Dimensionality and factor structure were explored through exploratory factor analysis, using a principal axis factoring method with oblimin rotation. The number of extracted factors was determined according to eigenvalues (>1) and visual inspection of the screeplot, taking the "elbow" as the point of separation.[29] All statistical analyses were carried out using STATA 12 for Windows (StataCorp, College Station, TX, USA).

Results

Table 1 presents descriptive statistics of sociodemographic data and applied rating scales. The mean age of the participants was 74.2 years (standard deviation [SD] 6.6), and 57.0% were women. Mean 10-item Center for Epidemiologic Studies Depression Scale score was 2.0 (SD 2.3) and, according to this scale, depression was present in 30.3% of the participants who answered the self-administered questionnaire. More than half of the participants (50.9%, $n = 360$) reported having three or more chronic conditions, the average number of conditions being 3.2 (SD 2.4). Mean DBMA score was 6.8 (SD 7.1).

Table 1 Characteristics of the study sample ($n = 707$)

Characteristic (range)		*n* (%)
		Mean ± SD
Sex	Male	304 (43.0)
	Female	403 (57.0)
Age (years)		74.2 ± 6.6
Education	Less than primary	365 (51.6)
	Primary	137 (19.4)
	Secondary	83 (11.7)
	University	122 (17.3)
Living area	<10 000 inhabitants	133 (18.8)
	10 000–100 000 inhabitants	273 (38.6)
	100 000–500 000 inhabitants	201 (28.4)
	>500 000 inhabitants	100 (14.1)
Marital status	Single	31 (4.4)
	Married/living with partner	439 (62.1)
	Widowed	211 (29.8)
	Divorced/separated	26 (3.7)
CES-D†	Depression	159 (22.5)
	No depression	366 (51.8)
	Missing	182 (25.7)
Physical functioning (24–96)		88.3 ± 11.8
PWI (0–100)		75.3 ± 11.1
Satisfaction with health (0–10)		7.1 ± 5.3
Self-reported number of conditions (0–21)		3.2 ± 2.4
DBMA (0–105)		6.8 ± 7.1

†Cut-off point: 3 out of 10. CES-D, Center for Epidemiologic Studies Depression Scale; DBMA, Disease Burden Morbidity Assessment; PWI, Personal Wellbeing Index; SD, standard deviation.

Hypertension was the most frequent chronic health condition ($n = 339$, 48.1%), and also the condition with the lowest mean disease burden score (mean score 1.5, SD 0.9; Table 2). Parkinson's disease was the least frequent condition ($n = 12$, 1.7%). The highest mean disease burden score was found for chronic back pain (2.9, SD 1.2) and myocardial infarction (2.9, SD 1.2).

All items had less than 4% missing responses, and the percentage of computable scores for the total scale was 88.4%. The observed and possible range was 0–5 for all items, with a median score of 0. The median for the total scale was 5, with a mean–median difference of 1.7%. For all items, floor effects were above 50% and ceiling effects were below 3%. When only studying the present conditions, there was still a floor effect, but less pronounced (range 14.1–66.8%). Skewness was 1.80.

Item-total corrected correlation was low for all conditions (range 0.10–0.49). Cronbach's alpha was 0.72, and the item homogeneity index was 0.09.

© 2016 Japan Geriatrics Society

Table 2 Prevalence of conditions, disease burden per condition and scale validation data in the analyzed sample ($n = 707$)

Medical condition		Self-reported conditions		Self-reported disease burden	DBMA scores used for scale validation†					
		Missing (%)	Condition present (%)	Mean (SD)	Missing (%)	Mean (SD)	Observed range	Floor Effect (%)	Ceiling Effect (%)	ITCC
1	Hypertension	2 (0.3)	339 (48.1)	1.5 (0.9)	19 (2.7)	0.7 (1.0)	0–5	53.2	0.3	0.24
2	Osteoarthritis	2 (0.3)	328 (46.5)	2.7 (1.2)	22 (3.1)	1.2 (1.6)	0–5	55.0	2.9	0.43
3	Circulation problems/ intermittent claudication	4 (0.6)	159 (22.6)	2.2 (1.2)	13 (1.8)	0.5 (1.1)	0–5	78.4	0.6	0.43
4	Rheumatoid arthritis	7 (1.0)	155 (22.1)	2.8 (1.2)	17 (2.4)	0.6 (1.3)	0–5	79.0	1.7	0.49
5	Chronic back pain	3 (0.4)	155 (22.0)	2.9 (1.2)	16 (2.3)	0.6 (1.3)	0–5	79.5	1.9	0.45
6	Depression	5 (0.7)	130 (18.5)	2.4 (1.3)	12 (1.7)	0.4 (1.1)	0–5	82.3	1.0	0.45
7	Urinary tract problems (prostate, bladder)	3 (0.4)	127 (18.0)	2.2 (1.2)	11 (1.6)	0.4 (1.0)	0–5	82.9	0.4	0.29
8	Osteoporosis	3 (0.4)	119 (16.9)	2.4 (1.3)	9 (1.3)	0.4 (1.0)	0–5	83.8	1.3	0.39
9	Diabetes	3 (0.4)	109 (15.5)	2.0 (1.1)	13 (1.8)	0.3 (0.8)	0–5	85.7	0.6	0.27
10	Anxiety	4 (0.6)	92 (13.1)	2.5 (1.3)	14 (2.0)	0.3 (0.9)	0–5	88.2	0.7	0.44
11	Cancer	11 (1.6)	76 (10.9)	2.0 (1.2)	12 (1.7)	0.2 (0.7)	0–5	89.2	0.1	0.13
12	Gastric/duodenal ulcer	6 (0.8)	74 (10.6)	1.7 (1.1)	8 (1.1)	0.2 (0.6)	0–5	89.7	0.3	0.17
13	Heart failure	4 (0.6)	74 (10.5)	2.4 (1.4)	14 (2.0)	0.2 (0.8)	0–5	90.8	0.9	0.15
14	Kidney disease	6 (0.8)	60 (8.6)	1.7 (1.1)	10 (1.4)	0.1 (0.5)	0–5	92.0	0.1	0.19
15	COPD/emphysema	5 (0.7)	56 (8.0)	2.5 (1.4)	6 (0.8)	0.2 (0.1)	0–5	92.2	0.9	0.19
16	Asthma	6 (0.8)	41 (5.9)	2.6 (1.3)	8 (1.1)	0.1 (0.7)	0–5	94.4	0.3	0.23
17	Angina	5 (0.7)	36 (5.1)	2.0 (1.2)	9 (1.3)	0.1 (0.5)	0–5	95.4	0.1	0.16
18	Myocardial infarction	5 (0.7)	32 (4.6)	2.9 (1.2)	9 (1.3)	0.1 (0.6)	0–5	96.0	0.3	0.10
19	Cerebral embolism/stroke	4 (0.6)	29 (4.1)	2.0 (1.4)	7 (1.0)	0.1 (0.5)	0–5	96.3	0.3	0.16
20	Memory disorders	4 (0.6)	26 (3.7)	2.6 (1.4)	11 (1.6)	0.1 (0.5)	0–5	97.3	0.4	0.17
21	Parkinson's disease	3 (0.4)	12 (1.7)	2.7 (1.3)	5 (0.7)	0.0 (0.3)	0–5	98.6	0	0.16
	Total				82 (11.6)	6.8 (7.1)	0–41	10.6	0	

†Self-reported disease burden scores including value 0 if condition not present. Items are presented in descending order of prevalence. COPD, chronic obstructive pulmonary disease; DBMA, Disease Burden Morbidity Assessment; ITCC, item-total corrected correlation; SD, standard deviation.

© 2016 Japan Geriatrics Society

Data on convergent validity are shown in Table 3. The DBMA had a correlation of –0.56 with physical functioning and perceived health, –0.41 with PWI and 0.41 with depression. All correlations were significantly stronger for the DBMA than for the number of diseases. Women had higher mean DBMA scores than men (8.4 *vs* 4.9, $P < 0.001$). Disease burden scores increased significantly with age, with a mean score of 6.1 in participants aged <75 years and 7.7 in participants aged ≥75 years ($P < 0.001$). The exploratory factor analysis extracted five factors (explained variance 43.6%): conditions of the locomotor system (intermittent claudication, osteoarthritis, rheumatoid arthritis, osteoporosis and chronic back pain), depression/anxiety, cardiovascular diseases (myocardial infarction, heart failure, angina, stroke), lung disorders (asthma, COPD), and a mixed group of cancer and renal/urinary tract diseases (kidney disease, cancer, urinary tract problems). Five conditions did not fit in any of the factors: hypertension, diabetes, gastric/duodenal ulcer, memory disorders and Parkinson's disease.

Discussion

The present study analyzed the psychometric properties of a Spanish version of the DBMA, and, despite some limitations, it was found to be a valid and reliable tool for measuring self-reported disease burden in older adults. Previous studies had already carried out initial validations of the DBMA. Bayliss *et al.* showed that the DBMA had stronger correlations with QoL outcomes than two other comorbidity measures, and found that median sensitivity and specificity relative to chart review were 75% and 92%, respectively.[9] Poitras *et al.* validated the DBMA in French in a general population aged 18 years and older, and found moderate-to-large correlations with the cumulative illness rating scale and a high test–retest reliability.[13]

As the DBMA repeats the same question and only for the conditions a participant has, it is easier and less time-consuming than other psychometric questionnaires. However, this design also meant that some of the present results did not meet the previously set criteria. Although it does assess a psychometric construct; that is, the disease burden that people experience as a result of their chronic conditions, the 21 items in the DBMA were not designed to be highly associated to each other, nor to the total score. Therefore, it was not a surprise to find a very low homogeneity index, and low item-total correlations. However, internal consistency was satisfactory.

In addition, because of the low prevalence of most of the conditions included in this questionnaire, a high value for skewness and large floor effects were not surprising. Floor effects were lower when only studying present conditions, but remained high. Reformulating the response scale options might attenuate this effect, and Rasch analysis could be helpful for this purpose. The percentage of computable scores was just below standards. The DBMA in the present study was part of a much larger questionnaire, which might have influenced the willingness and capability to answer the questions of the DBMA.

The DBMA was found to have a strong correlation with physical functioning and perceived health, and moderate correlations with depression and PWI. Bayliss *et al.* reported similar associations, with comparable correlations with physical functioning and self-reported health status, and a lower correlation with depression.[9,11,12] This last difference can be explained by the fact that, unlike Bayliss *et al.*, we did include depression in the list of studied conditions. Depression is a frequent condition and an important cause of disability,[30,31] for which we considered it essential to be included in the DBMA, as was also done by Poitras *et al.*[13]

ELES-PS data confirmed that the DBMA is more strongly associated with depression, perceived health and physical functioning than the number of conditions, although these differences were less pronounced than in the study published by Bayliss *et al.*[9] This outcome is of interest because although the DBMA is a relatively short questionnaire, it is still more laborious than a simple disease count. The association between a simple count of chronic conditions as a measure of multimorbidity and QoL has widely been studied.[32,33] However, measures taking into account disease severity seem to be better predictors of QoL, as was confirmed in the present study.[34] The relationship between the DBMA and the PWI had not been studied yet, and the present study has found a moderate correlation. Some of the dimensions of this scale are not directly related to health, such as personal safety and future security,[20] and the correlation with

Table 3 Convergent validity: Spearman's rank correlation coefficients of the self-reported number of conditions and Disease Burden Morbidity Assessment with other health-related measurements

	No. conditions	DBMA	*P*-level for correlation difference
Depression, CES-D ($n = 474$)	0.35*	0.41*	0.0043
Physical functioning ($n = 596$)	–0.51*	–0.56*	0.0010
Quality of Life, PWI ($n = 540$)	–0.35*	–0.41*	0.0006
Perceived health ($n = 625$)	–0.51*	–0.56*	0.0035

*$P < 0.001$. CES-D, Center for Epidemiologic Studies Depression Scale; DBMA, Disease Burden Morbidity Assessment; PWI, Personal Wellbeing Index.

© 2016 Japan Geriatrics Society

a scale measuring health-related QoL could possibly be stronger.

The factor analysis identified known disease groups. Intermittent claudication is a motor symptom and therefore it was probably more related to musculoskeletal disorders than to cardiovascular diseases.

Some limitations should be acknowledged. As the sample used for this validation study contained a subsample from the Basque Country, results cannot be extrapolated to the Spanish population as a whole. Weighing the sample to correct for this overrepresentation would have been the most appropriate solution, but impeded some of the feasibility and acceptability analyses. Also, as no other comorbidity indexes were included in the ELES-PS, we could not analyze convergent validity with other more widely used comorbidity tools. Furthermore, as in other studies, we adapted the list of conditions included in the DBMA, which hinders comparisons across studies.[12–14] A standard list of conditions should be developed for future use.

In conclusion, the DBMA has shown to be a valid and easy-to-use instrument for measuring disease burden. As this scale provides a subjective measure of multimorbidity, it is particularly useful for investigations using QoL outcomes. Future research should include a Rasch analysis in order to further study the scale's dimensionality and to explore other measurement properties.

Acknowledgment

The ELES-PS was supported by the Spanish Ministry of Science and Innovation, National Plan of Research and Development (ref. CSO2011-30210-C02-01).

This work was partially supported by ENCAGE-CM (ref: S2015/HUM-3367), funded by the I+D Activity Program of Madrid Community research groups on social sciences and humanities and co-funded by the European Social Fund.

This work also arises from the Joint Action on Chronic Diseases and Promoting Healthy Ageing across the Life Cycle (JA-CHRODIS), which has received funding from the European Union, in the framework of the Health Programme (2008–2013). Sole responsibility lies with the author and the Consumers, Health, Agriculture and Food Executive Agency is not responsible for any use that may be made of the information contained therein.

I Wijers designed and executed the statistical analysis, interpreted the data and prepared the manuscript. A Ayala assisted with the statistical analysis and reviewed the manuscript for important conceptual content. A Rodriguez-Laso and V Rodriguez-Rodriguez designed the study, supervised the acquisition of data and reviewed the manuscript for important conceptual content. C Rodriguez-Blazquez reviewed the manuscript for important conceptual content. MJ Forjaz assisted with writing and reviewed the manuscript for important conceptual content. All authors approved the final version.

Disclosure statement

The authors declare no conflict of interest.

References

1 Gijsen R, Hoeymans N, Schellevis FG, Ruwaard D, Satariano WA, van den Bos GA. Causes and consequences of comorbidity: a review. *J Clin Epidemiol* 2001; **54**: 661–674.
2 Newman AB. Comorbidity and Multimorbidity. In: Newman AB, Cauley JA, eds. *The Epidemiology of Aging*. Netherlands: Springer, 2012; 119–133.
3 Marengoni A, Angleman S, Melis R *et al.* Aging with multimorbidity: A systematic review of the literature. *Ageing Res Rev* 2011; **10**: 430–439.
4 Sangha O, Stucki G, Liang MH, Fossel AH, Katz JN. The Self-Administered Comorbidity Questionnaire: a new method to assess comorbidity for clinical and health services research. *Arthritis Rheum* 2003 (Apr 15); **49** (2): 156–163.
5 Forjaz MJ, Rodriguez-Blazquez C, Ayala A *et al.* Chronic conditions, disability, and quality of life in older adults with multimorbidity in Spain. *Eur J Intern Med* 2015; **26**: 176–181.
6 Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987; **40**: 373–383.
7 Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. *Med Care* 1998; **36**: 8–27.
8 Crabtree HL, Gray CS, Hildreth AJ, O'Connell JE, Brown J. The Comorbidity Symptom Scale: a combined disease inventory and assessment of symptom severity. *J Am Geriatr Soc* 2000; **48**: 1674–1678.
9 Bayliss EA, Ellis JL, Steiner JF. Subjective assessments of comorbidity correlate with quality of life health outcomes: initial validation of a comorbidity assessment instrument. *Health Qual Life Outcomes* 2005; **3**: 51.
10 Bayliss EA, Ellis JL, Shoup JA, Zeng C, McQuillan DB, Steiner JF. Association of patient-centered outcomes with patient-reported and ICD-9-based morbidity measures. *Ann Fam Med* 2012; **10**: 126–133.
11 Bayliss EA, Ellis JL, Steiner JF. Barriers to Self-Management and Quality-of-Life Outcomes in Seniors With Multimorbidities. *Ann Fam Med* 2007; **5**: 395–402.
12 Bayliss E, Ellis J, Steiner J. Seniors' self-reported multimorbidity captured biopsychosocial factors not incorporated in two other data-based morbidity measures. *J Clin Epidemiol* 2009; **62**: 550–7.e1.
13 Poitras M-E, Fortin M, Hudon C, Haggerty J, Almirall J. Validation of the disease burden morbidity assessment by self-report in a French-speaking population. *BMC Health Serv Res* 2012; **12**: 35.
14 Hudon C, Fortin M, Poitras M-E, Almirall J. The relationship between literacy and multimorbidity in a primary care setting. *BMC Fam Pract* 2012 (Apr 26); **13** (1): 33.
15 Byles JE, D'Este C, Parkinson L, O'Connell R, Treloar C. Single index of multimorbidity did not predict multiple outcomes. *J Clin Epidemiol* 2005; **58**: 997–1005.
16 Fried LP, Bandeen-Roche K, Kasper JD, Guralnik JM. Association of comorbidity with disability in older women: the Women's Health and Aging Study. *J Clin Epidemiol* 1999; **52**: 27–37.

© 2016 Japan Geriatrics Society

17 Groll DL, To T, Bombardier C, Wright JG. The development of a comorbidity index with physical function as the outcome. *J Clin Epidemiol* 2005; **58**: 595–602.
18 Robison J, Gruman C, Gaztambide S, Blank K. Screening for depression in middle-aged and older puerto rican primary care patients. *J Gerontol A Biol Sci Med Sci* 2002; **57**: M308–M314.
19 Health and Retirement Study [Internet]. [Cited 27 Oct 2014]. Available from: http://hrsonline.isr.umich.edu/index.php
20 The International Wellbeing Group. Personal Wellbeing Index-Adult (PWI-A). Manual 2013. [Internet]. 5th ed. Melbourne: Australian Centre on Quality of Life, Deakin University; 2006 [Cited 11 Dec 2013]. 41p p. Available from: http://www.deakin.edu.au/research/acqol/instruments/wellbeing-index/pwi-a-english.pdf
21 Forjaz MJ, Ayala A, Rodriguez-Blazquez C *et al.* Rasch analysis of the International Wellbeing Index in older adults. *Int Psychogeriatr* 2012; **24**: 324–332.
22 Martinez-Martin P, Forjaz MJ. How to Evaluate Validation Data. In: Sampaio C, Goetz CG, Schrag A, eds. *Rating Scales in Parkinson's Disease*. Oxford University Press, 2012; 16–41.
23 Virués-Ortega J, Rodríguez-Blázquez C, Micheli F, Carod-Artal FJ, Serrano-Dueñas M, Martínez-Martín P. Cross-cultural evaluation of the modified Parkinson Psychosis Rating Scale across disease stages. *Mov Disord* 2010 (Jul 30); **25** (10): 1391–1398.
24 Diaz-Redondo A, Rodriguez-Blazquez C, Ayala A, Martinez-Martin P, Forjaz MJ, Spanish Research Group on Quality of Life and Aging. EQ-5D rated by proxy in institutionalized older adults with dementia: psychometric pros and cons. *Geriatr Gerontol Int* 2014; **14**: 346–353.
25 Caughey GE, Vitry AI, Gilbert AL, Roughead EE. Prevalence of comorbidity of chronic diseases in Australia. *BMC Public Health* 2008 (Jun 27); **8** (1): 221.
26 Cohen J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd edn. Hillsdale, N.J: Routledge, 1988.
27 Diedenhofen B, Musch J. cocor: A Comprehensive Solution for the Statistical Comparison of Correlations. *PLoS One* 2015 (Apr 2); **10** (4): e0121945.
28 Violan C, Foguet-Boreu Q, Flores-Mateo G *et al.* Prevalence, Determinants and Patterns of Multimorbidity in Primary Care: A Systematic Review of Observational Studies. *PLoS One* 2014 (Jul 21); **9** (7): e102149.
29 Henry C, M'Bailara K, Mathieu F, Poinsot R, Falissard B. Construction and validation of a dimensional scale exploring mood disorders: MAThyS (Multidimensional Assessment of Thymic States). *BMC Psychiatry* 2008; **8**: 82.
30 Moussavi S, Chatterji S, Verdes E, Tandon A, Patel V, Ustun B. Depression, chronic diseases, and decrements in health: results from the World Health Surveys. *The Lancet* 2007 (Sep 14); **370** (9590): 851–858.
31 Ferrari AJ, Charlson FJ, Norman RE *et al.* Burden of depressive disorders by country, sex, age, and year: findings from the global burden of disease study 2010. *PLoS Med* 2013; **10**: e1001547.
32 Wensing M, Vingerhoets E, Grol R. Functional status, health problems, age and comorbidity in primary care patients. *Qual Life Res* 2001; **10**: 141–148.
33 Cheng L, Cumber S, Dumas C, Winter R, Nguyen KM, Nieman LZ. Health related quality of life in pregeriatric patients with chronic diseases at urban, public supported clinics. *Health Qual Life Outcomes* 2003; **1**: 63.
34 Fortin M, Hudon C, Dubois M-F, Almirall J, Lapointe L, Soubhi H. Comparative assessment of three different indices of multimorbidity for studies on health-related quality of life. *Health Qual Life Outcomes* 2005; **3**: 74.

© 2016 Japan Geriatrics Society

Copyedited by: AM

THE GERONTOLOGICAL SOCIETY OF AMERICA®

The Gerontologist
cite as: *Gerontologist*, 2017, Vol. 00, No. 00, 1–9
doi:10.1093/geront/gnx061
Advance Access publication XXXX XX, XXXX

OXFORD

Measurement Online

Rasch Analysis and Construct Validity of the Disease Burden Morbidity Assessment in Older Adults

Irene Geurtje Maria Wijers, MSc,[1] Alba Ayala, MSc,[2] Carmen Rodriguez-Blazquez, PhD,[3] Angel Rodriguez-Laso, PhD,[4] Vicente Rodriguez-Rodriguez, PhD,[5] and Maria João Forjaz, PhD[2,*]

[1]Department of Preventive Medicine and Quality Management, General University Hospital Gregorio Marañón, Madrid, Spain. [2]National School of Public Health, Institute of Health Carlos III and REDISSEC, Madrid, Spain. [3]National Centre of Epidemiology, Carlos III Institute of Health and CIBERNED, Madrid, Spain. [4]Biomedical Research Foundation of the University Hospital of Getafe, Madrid, Spain. [5]Institute of Economics, Geography and Demography, Spanish National Research Council, Madrid, Spain.

*Address correspondence to: Maria João Forjaz, PhD, Escuela Nacional de Sanidad/National School of Public Health, Epidemiology and Biostatistics Department, Avda. Monforte de Lemos 5, Pab. 7, Despacho 07.02.0019, 28029 Madrid, España/Spain. E-mail: jforjaz@isciii.es

Received: September 12, 2016; Editorial Decision Date: April 12, 2017

Decision Editor: Rachel Pruchno, PhD

Abstract

Purpose of the Study: The Disease Burden Morbidity Assessment (DBMA) is a self-report questionnaire in which participants rate the disease burden caused by a number of medical conditions. This paper studies the measurement properties of the DBMA, using Rasch analysis.

Design and Methods: We used data of 1,400 community-dwelling adults aged 50 years and older participating in the Ageing in Spain Longitudinal Study, Pilot Survey (ELES-PS). Test of fit to the Rasch model, reliability, unidimensionality, response dependency, category structure, scale targeting, and differential item functioning (DIF) were studied in an iterative way. Construct validity of the linear measure provided by the Rasch analysis was subsequently assessed.

Results: To achieve an adequate fit to the Rasch model, all items were rescored by collapsing response categories. Reliability (Person Separation Index) was low. The scale was unidimensional and neither response dependency nor relevant DIF were found. The linear measure had a correlation of –0.48 with physical functioning, –0.47 with perceived health, 0.32 with depression, and –0.24 with quality of life (QoL) and displayed satisfactory known-groups validity by sex and age groups. Relative precision analysis showed that the linear measure discriminated better between age groups than the original raw score, but for sex no difference was found.

Implications: Despite some limitations, support was found for the validity of the DBMA in older adults. Its linear scores may be useful to assess strategies aimed at improving the QoL of patients with multimorbidity. More research is needed in a hospital-based sample.

Keywords: Burden of Illness, Chronic disease, Comorbidity

The ageing of populations is a global phenomenon (Beard, Officer, & Cassels, 2016). In 1950, the worldwide proportion of persons aged 60 years and older was 8% (United Nations, Department of Economic and Social Affairs, Population Division, 2013). This percentage rose to almost 12% in the year 2013, and is expected to reach 21% in the year 2050. In high-income countries, these rates are even higher, with almost 23% in the year 2013 and a prediction of 32% for the year 2050. As a consequence, the prevalence of multimorbidity, that is, the presence of multiple

© The Author 2017. Published by Oxford University Press on behalf of The Gerontological Society of America. All rights reserved.
For permissions, please e-mail: journals.permissions@oup.com.

Copyedited by: AM

coexisting chronic conditions in one person, will also continue to increase (Gijsen et al., 2001). Multimorbidity is a worldwide health problem with well-described associations with mortality, complications of treatment, health care utilization, and a negative effect on quality of life (QoL) (Barnett et al., 2012; Fortin et al., 2004; Gijsen et al., 2001; Sangha, Stucki, Liang, Fossel, & Katz, 2003). It requires a different health care approach, with a more holistic view of patients instead of treating single diseases (Fortin, Stewart, Poitras, Almirall, & Maddocks, 2012).

There are different tools to assess multimorbidity, and the choice of instrument depends on the methodology and outcomes of the investigation (Bayliss, Ellis, & Steiner, 2009). Several studies have shown the importance of assessing subjective disease severity when studying multimorbidity, especially in relation to QoL outcomes (Byles, D'Este, Parkinson, O'Connell, & Treloar, 2005; Crabtree, Gray, Hildreth, O'Connell, & Brown, 2000; Sangha et al., 2003). Therefore, Bayliss, Ellis, and Steiner (2005) created a patient-reported outcome measure in which patients select chronic conditions from a list and then rate their impact on everyday activities as a measure of disease severity. This was conceptualized as self-reported disease burden. Disease burden can be defined as the impact of disease events on various dimensions of human life (Pinheiro, Plaß, & Krämer, 2011), in this case the subjective interference with daily activities. The tool was subsequently denominated the Disease Burden Morbidity Assessment (DBMA) by Poitras, Fortin, Hudon, Haggerty, and Almirall (2012).

After an initial validation according to the classical test theory, including an exploratory factor analysis (Wijers et al., 2016), an additional step in the validation process would be an analysis following the item response theory. The aim of this study was to perform this through a Rasch analysis (Rasch, 1980). Rasch analysis allows us to study scale attributes such as unidimensionality, response category ordering, local independence of items, item bias by specific groups, and scale targeting, and provides a linear measure, which, given an appropriate distribution, permits the use of parametric statistics (Forjaz et al., 2012).

Methods

Study Design and Sample

Data came from the Ageing in Spain Longitudinal Study, Pilot Survey (ELES-PS), which included 1,747 community-dwelling adults aged 50 or more, living in Spain (Teófilo Rodríguez, González Cabezas, Díaz Veiga, & Rodríguez Rodríguez, 2011). For the sampling, stratified clusters of census sections were randomly selected by autonomous region and municipality, proportionally to their population of 50 years and older. Households with a telephone line were selected at random from a commercial household telephone directory. Per household, individuals aged 50 or more were randomly selected, with post-stratification by sex and age group (50–59, 60–69, 70–79, and 80–89 years). Field work was conducted in 2011.

The data in the ELES-PS study were collected in four stages: a telephone questionnaire (n = 1,747), a visit by a trained nurse (n = 1,531), a Computer-Assisted Personal Interviewing (CAPI) questionnaire (n = 1,400), and a self-administered questionnaire (n = 1,145). DBMA data were collected through the CAPI questionnaire, and its 1,400 participants formed the sample that was used for the current study. For the Rasch analysis, a random subsample of 300 was taken, since analysis with samples larger than 300 could result in statistically significant deviations from the Rasch model of otherwise well-fitting items (Linacre, 1994, 2016; Mavranezouli, Brazier, Young, & Barkham, 2011; Smith, Rush, Fallowfield, Velikova, & Sharpe, 2008).

Descriptive statistics of sociodemographic data and applied rating scales of the total sample and the subsample are displayed in Table 1. Mean age of the participants in the total sample was 65.5 (standard deviation [*SD*] = 10.40) years, and 55.36% of them were women. The mean number of self-reported diseases was 2.5 (*SD* = 2.25), and a mean raw DBMA score of 5.29 (*SD* = 6.39) was found.

Assessments

The DBMA, first described by Bayliss and colleagues (2005), consists of a self-report questionnaire in which participants rate the disease burden caused by a number of medical conditions, if present. Patients are asked to what extent conditions interfere with daily activities, on a 5-point scale from 1 (not at all) to 5 (a lot). Conditions not present are scored zero. We adapted the list of conditions included by Bayliss and colleagues by selecting 21 common chronic conditions, according to their use in other multimorbidity indexes (Bayliss et al., 2005; Byles et al., 2005; Fried, Bandeen-Roche, Kasper, & Guralnik, 1999; Groll, To, Bombardier, & Wright, 2005; Sangha et al., 2003). More detailed information about how the 21 included conditions were selected may be found elsewhere (Wijers et al., 2016).

To measure physical functioning, a 24-item list of different basic and instrumental activities of daily living, as used in the Health and Retirement Study (Bendayan et al., 2016), was applied. Participants were asked whether they experience difficulties when performing these activities on a scale from 1 (always) to 4 (never). Higher scores indicate better physical functioning.

A dichotomous, self-administered 10-item Center for Epidemiological Studies-Depression (CES-D) scale was used to screen for depression cases (scores of 3+) (Robison, Gruman, Gaztambide, & Blank, 2002). Previous studies found support for this short version of the CES-D to be as reliable as the original CES-D, with a Cronbach's alpha of 0.80, and to show satisfactory convergent validity with the Composite International Diagnostic Interview (sensitivity and specificity of 84% and 64%, respectively) (Irwin, Artin, & Oxman, 1999; Robison et al., 2002).

For QoL, the CAPI questionnaire contained the Personal Wellbeing Index (PWI) (The International Wellbeing Group,

Copyedited by: AM

Table 1. Characteristics of the Study Sample (n = 1,400) and Rasch Analysis Subsample (n = 300)

Characteristic (range)	Total sample *N* (%) Mean ± *SD*	Subsample *N* (%) Mean ± *SD*
Sex		
Men	625 (44.64)	132 (44.00)
Women	775 (55.36)	168 (56.00)
Age (years)	65.50 ± 10.40	64.96 ± 10.27
Education		
Less than primary	480 (34.30)	98 (32.67)
Primary	313 (22.40)	77 (25.67)
Secondary	298 (21.30)	61 (20.33)
University	309 (22.10)	64 (21.33)
Living area		
<10.000 inhabitants	315 (22.50)	70 (23.33)
10.000–100.000 inhabitants	502 (35.90)	103 (34.33)
100.000–500.000 inhabitants	385 (27.50)	90 (30.33)
>500.000 inhabitants	198 (14.10)	37 (12.33)
Marital status		
Single	75 (5.36)	17 (5.67)
Married/living with partner	1014 (72.43)	225 (75.00)
Widowed	244 (17.43)	44 (14.67)
Divorced/separated	67 (4.80)	14 (4.67)
CES-D[a]		
Depression	297 (21.21)	78 (26.00)
No depression	797 (56.93)	160 (53.33)
Missing	306 (21.86)	62 (20.67)
PWI (0–100)	74.94 ± 11.09	73.34 ± 12.52
Satisfaction with health (0–10)	7.19 ± 3.92	6.84 ± 1.91
Physical functioning (24–96)	91.06 ± 9.56	90.65 ± 10.10
Number of self-reported conditions (0–21)	2.50 ± 2.25	2.61 ± 2.29
DBMA raw score (0–105)	5.29 ± 6.39	5.36 ± 6.19

Note: CES-D = Center for Epidemiological Studies-Depression scale; DBMA = Disease Burden Morbidity Assessment; PWI = Personal Wellbeing Index; *SD* = standard deviation.
[a]Cutoff point: 3 out of 10.

2013). This scale consists of seven life dimensions rated on a 1–10 scale. Total scores were linearly transformed into a 0–100 scale, higher scores indicating better QoL. Previous research found support for the validity and reliability of this linear measure in older adults, correlating moderately with "satisfaction with life" and showing a PSI of 0.91 (Forjaz et al., 2012). The dimension "personal health" of the PWI was used as a measure of perceived health.

Statistical Analysis

Rasch analysis was performed using RUMM 2030. Differences between thresholds were not expected to be equal across items, so the Masters Partial Credit polytomous model was chosen (Masters, 1982), which was confirmed by a significant likelihood ratio statistic. Test of fit to the Rasch model, reliability, unidimensionality, response dependency, category structure, scale targeting, and differential item functioning (DIF) were studied in an iterative way, making model modifications until a good fit was achieved (Tennant & Conaghan, 2007).

Fit to the Rasch model was tested by comparing the observed data with the theoretical item performance according to the Rasch model. The item–trait interaction statistic, reported as a chi-square, needs to be nonsignificant (Tennant & Conaghan, 2007). Item and person summary fit statistics should follow a normal distribution with a mean and *SD* of approximately 0 and 1, respectively. Individual item and person standardize fit residuals should be within the ±2.5 range and chi-square differences for items and persons should be nonsignificant with Bonferroni correction for number of items (Pallant & Tennant, 2007).

Reliability was determined with the Person Separation Index (PSI), which is interpreted similarly to Cronbach's coefficient alpha: a minimum value of 0.70 for group comparisons is recommended (Tennant & Conaghan, 2007). The PSI was also obtained in RUMM2020 since algorithms derived from this program provide reliability results less influenced by extreme values, missing data, and floor and ceiling effects than those obtained with RUMM2030 (Forjaz et al., 2015).

Unidimensionality was tested through a Principal Component Analysis (PCA) of the residuals (Smith et al., 2006). This test defines two subsets of items, those positively and those negatively correlated with the first residual factor, and the difference in these estimates for each person are compared with a t test. The percentage of significant t tests should not exceed 5% (Tennant & Conaghan, 2007).

Response dependency was assessed through the residual correlation index and a correlation of >0.30 was taken as an indication of local dependency (Forjaz et al., 2012). Category structure was explored through category probability curves, and in case of disordered thresholds, items were rescored by collapsing adjacent categories. Scale targeting was assessed through visual inspection of the person–item map, showing the distribution of persons and items along the construct.

DIF examines whether different groups within the sample, despite of equal levels of the characteristic being measured, respond in a different manner to an individual item (Tennant & Pallant, 2007). We studied DIF for age (<65 vs ≥65 years, which was the median value in our sample), sex, and educational level (primary school or less vs more than primary school). DIF analysis was performed through an analysis of variance (ANOVA) with Bonferroni correction. In case DIF was identified, this was further analyzed through a top-down purification approach. In this approach, items are divided into two groups, according to the presence of absence of DIF, and these are applied as

Copyedited by: AM

two testlets (or superitems). If the superitem formed by the items with DIF does not present DIF, then DIF is considered to cancel out (Tennant, Penta, et al., 2004).

Once fit to the Rasch model was achieved, disease burden scores of the total sample were used to calculate a linear measure, on a logit scale, which was converted into a 0–47 range through a linear transformation. In order to compare the subsample of 300 and the rest of the sample (n = 1,100), a paired-sample t test was done, comparing the logit estimation of the two samples (300 vs 1,100) for each raw score. Anchor values of the sample of 300 were used to fix item estimations of the other sample. In addition, a DIF analysis with the sample as a factor was performed.

Psychometric attributes of the linear measure according to the classical test theory were analyzed using Stata 12 version for Windows: mean-to-median differences (criterion, <10%), floor and ceiling effects (<15%), and skewness (−1 to 1) were calculated for acceptability (Virués-Ortega et al., 2010). Construct validity was assessed through known-groups validity for sex and age (<65 vs ≥ 65 years) and convergent validity with other health outcomes. We hypothesized to find higher disease burden scores for women (Barnett et al., 2012) and in the highest age group (Global Burden of Disease Study 2013 Collaborators, 2015), which was studied with a Wilcoxon rank-sum test due to the non-normal distribution of the linear measure. For convergent validity, we calculated Spearman's rank correlations with physical functioning, depression (CES-D total score), QoL (PWI), and perceived health (PWI item 2: personal health). Moderate to high correlations (r > 0.30) were expected (Cohen, 1988).

We performed a relative precision analysis in order to assess how much more or less precise the Rasch-based score is relative to the raw summative-based score in distinguishing groups expected to differ (Las Hayas et al., 2011). This was done for sex and age groups (<65 vs ≥ 65 years). Relative precision was calculated as the ratio of pairwise Z statistics (the linear measure Z statistic divided by the raw score Z statistic) (Sakthong, Charoenvisuthiwongs, & Shabunthom, 2008), and a bootstrap method was applied in order to obtain confidence intervals (CIs) for relative precision statistics (Fitzpatrick et al., 2004). Rasch analysis takes into account observations with missing values when calculating the linear measure. In order to be able to calculate the relative precision, the studied sample sizes should be equal, thus observations with missing values were excluded in the latter analysis.

Results

Rasch Analysis

The initial analysis, with the whole study sample, displayed poor fit to the Rasch model (Table 2). After selecting a subsample of 300, the fit indices improved, but still did not meet the fit criteria. Category probability curves showed disordered thresholds, so items were rescored to two (two items), three (13 items), four (five items), or five categories (one item) (Table 3). After this, the DBMA showed an acceptable fit to the Rasch model (Table 2). Individual item and person fit residuals were within the −2.5 to +2.5 range, with nonsignificant chi-squares (Table 3). However, PSI remained low, 0.272. When repeating this estimation in RUMM2020, the PSI improved to 0.637. In the PCA of the residuals, 0.72% of tests were outside the previously set range, indicating unidimensionality. All items were locally independent, with a residual correlation index ranging 0.000–0.188. No DIF was found for age or educational level. Four items showed DIF by sex of small magnitude (<0.5 logits): item 1 (hypertension), 14 (anxiety), 17 (osteoporosis), and 21 (urinary tract problems). In the top-down purification approach, this DIF was no longer present. The person–item threshold distribution (Figure 1) showed a floor effect and no persons represented the scale's higher levels of disease burden. Also, there were very few persons located around the highest point of the test information curve.

DBMA scores of the total sample were converted into a linear measure from 0 to 47 (see Supplementary Material). When comparing the subsample of 300 and the rest of the sample, no significant differences between the estimations were found (difference = 0.259 logits, t test = 1.226, p value = .226, and no DIF was observed by sample. AQ4

Classic Psychometric Analysis of the Linear Measure

Mean score of the linear measure was 7.36 (SD = 5.01), median score 7.44, with a mean–median difference of 0.17%. Floor effect for the total scale was 18.11%, with no ceiling effect, and skewness was 0.046. The linear measure presented a correlation of −0.48 with physical functioning, −0.47 with perceived health, 0.32 with depression (CES-D), and −0.24 with the PWI (p < .001). Women scored significantly higher than men, with mean scores of 8.14 (SD = 5.15) and 6.40 (SD = 4.65), respectively (p < .001), and scores increased with age: mean score among persons <65 years was 5.93 (SD = 4.76) versus a mean score of 8.77 (SD = 4.84) in persons aged ≥65 years (p < .001). The results of the relative precision analysis are shown in Table 4. The ability to discriminate between age groups increased by 9% when using the linear measure versus the raw score (95% CI: 1.03–1.17), but precision decreased 4% for age groups, although this difference was not statistically significant (95% CI: 0.86–1.05).

Discussion

This study analyzed the measurement properties of the DMBA according to the Rasch model. Rasch analysis provided knowledge of DBMA psychometric attributes that were not previously known. Test of fit to the Rasch model was satisfactory after rescoring response options,

Copyedited by: AM

Table 2. Global Fit to the Rasch Model of the DBMA Using the Total Sample (n = 1,400), After Selecting a Subsample (n = 300) and After Rescoring the Response Scale

		Standard	Total sample	Subsample	After rescoring
Item fit residual	Mean	0	−2.90	−1.12	−0.68
	SD	1	1.83	0.77	0.71
Person fit residual	Mean	0	−0.48	−0.42	−0.28
	SD	1	0.59	0.58	0.55
Item–trait interaction	χ^2	Low	316.33	165.14	154.43
	Prob.	NS	<.001	.89	.32
PSI		>0.70	0.07	0.14	0.27
Unidimensionality	Significant *t* tests (%)	<5	0.72	2.00	2.00

Note: DBMA = Disease Burden Morbidity Assessment; NS = nonsignificant; Prob. = probability; PSI = Person Separation Index; *SD* = standard deviation. Item fit residual refers to the difference between the data observed and the expected values at item level. Person fit refers to the difference between the data observed and the expected values at person level. Item–trait interaction is a chi-square value and probability resulting from the comparison between the expected and the mean observed score for groups of people with similar ability estimates. PSI is a reliability measure. Unidimensionality refers to the existence of one measurement construct (dimension) underlying the set of items.

Table 3. Threshold Ordering of Polytomous Items and Individual Item Fit to the Rasch Model After Rescoring (n = 300)

	Original categories										
	0	1	2	3	4	5	Individual item fit to the Rasch model				
Item	Rescored categories						Location	*SE*	Residual	χ^2	Prob.
Osteoarthritis	0	1	1	1	2	3	−1.914	0.093	−2.020	14.737	.040
Rheumatoid arthritis	0	1	1	1	2	3	−1.514	0.108	−1.749	5.988	.541
Chronic back pain	0	1	1	1	1	2	−1.392	0.130	−0.707	10.910	.143
Depression	0	1	1	1	1	2	−1.161	0.140	−0.996	7.631	.366
Circulation problems/intermittent claudication	0	1	1	1	2	3	−0.965	0.134	−1.698	5.237	.631
Hypertension	0	1	2	2	3	4	−0.934	0.088	−0.515	16.122	.024
Anxiety	0	1	1	1	1	2	−0.858	0.160	−1.474	7.290	.399
Osteoporosis	0	1	1	1	1	2	−0.644	0.162	−1.271	3.796	.803
Cancer	0	1	1	1	2	3	−0.515	0.160	0.068	8.254	.311
Diabetes	0	1	1	1	1	2	−0.371	0.174	−0.637	6.877	.442
Heart failure	0	1	1	1	1	2	−0.331	0.196	−0.055	8.192	.316
Urinary tract problems (prostate, bladder)	0	1	1	1	2	3	−0.219	0.147	−0.436	6.327	.502
COPD/emphysema	0	1	1	1	1	2	−0.137	0.222	−0.162	5.471	.603
Cerebral embolism/stroke	0	1	1	1	1	1	0.296	0.353	0.171	10.195	.178
Memory disorders	0	1	1	1	1	1	0.310	0.355	−0.424	5.818	.561
Gastric/duodenal ulcer	0	1	1	1	1	2	1.081	0.199	0.849	9.992	.189
Kidney disease	0	1	1	1	1	2	1.190	0.218	−0.444	7.298	.399
Asthma	0	1	1	1	1	2	1.578	0.274	−1.258	5.272	.627
Myocardial infarction	0	1	1	1	1	2	1.685	0.296	−0.340	2.781	.904
Angina	0	1	1	1	1	2	2.039	0.473	−0.312	4.173	.760
Parkinson's disease	0	1	1	2	2	2	2.776	0.924	−0.912	2.071	.956

Note: COPD = chronic obstructive pulmonary disease; Prob. = probability; *SE* = standard error. Items are ordered by increasing difficulty (mean location of thresholds).

with all items showing a good fit, and the scale was unidimensional. The residual correlation index did not identify response dependency, meaning that there were no items linked in such way that the response to one item would determine the response to another. Furthermore, Rasch analysis provided a linear measure, which allows calculation of change scores and, given a normal distribution, the use of parametric statistics (Hobart, Cano, Zajicek, & Thompson, 2007; Tennant, McKenna, & Hagell, 2004).

In order to achieve an adequate fit to the Rasch model, items needed to be rescored. This might have been due to too many response categories, which could have prevented people from making fine distinctions between rating scale steps. In most cases, response options were reduced to three categories. This reduction in response categories does not

Copyedited by: AM

require changing the original questionnaire. Instead, it may be performed when calculating the total scores, thus avoiding using different response categories that could be confusing to the respondent. It would be interesting to study whether simplifying the questionnaire, by reducing the response categories in the same way for all items, would improve the psychometric properties of the DBMA. However, this might reduce the scale precision.

Some items displayed DIF by sex, indicating that men and women, despite having the same level of burden caused by hypertension, anxiety, osteoporosis, or urinary tract problems, answered differently to these items. A very strict approach would have been to delete these items; however, this would have compromised the clinical applicability of the scale. Another possibility would be to split the items and get different calibrations for men and women (Tennant & Conaghan, 2007). This would make the scale more difficult to score, which, nonetheless, is justifiable if DIF results are replicated in further studies. DIF was no longer present in the top-down purification analysis, meaning that if DIF favors men for one item, to balance women are favored for another item. So, for the moment, and taking into consideration that differences were of small magnitude, we decided to be conservative and avoid scale modifications due to DIF. We do not expect DIF to have influenced the sex differences found in this study, since DIF refers to group differences at the same construct level.

The high floor effect (and as a consequence, the asymmetrical person–item threshold distribution) represents cases in which participants reported not having certain conditions. This implies that the floor effect can actually be regarded as an indicator of how "healthy" the studied population is. Also, there were very few persons located near the highest point of the test information curve, which represents the location where the test is the most powerful in the sense of measurement precision. These data suggest that the test performance would probably improve in a hospital-based sample, with a higher proportion of multimorbid patients and therefore less floor effects and better scale targeting; thus, more research is needed.

We found a low reliability in RUMM2030 and although the PSI value improved when using RUMM2020, it still did not fit the criterion. This effect is probably due to the design of the DBMA, in which disease burden is rated for single diseases. Experiencing disease burden from one disease does not imply that a person should have the other 20 conditions as well, which makes items in this scale less related to each other than in other scales. Nevertheless, having comorbid conditions does increase the disease burden experienced from a specific disease (Gadermann, Alonso, Vilagut, Zaslavsky, & Kessler, 2012; Moussavi et al., 2007), which might have caused that PSI in RUMM2020, less influenced by floor effects than in RUMM2030, was closer to the criterion of >0.70. This could also make us expect the PSI to be higher in a sample with more multimorbidity. Low PSIs were also found in other widely used scales, such as the EQ-5D-3D (Pickard, De Leon, Kohlmann, Cella, & Rosenbloom, 2007).

Our study showed moderate correlations between the linear measure and physical functioning, perceived health, and depression and a weak correlation with QoL. Bayliss and colleagues (2005) reported high correlations with physical functioning and perceived health (−0.63 and 0.60, respectively) and a weak correlation (−0.29) with depression. The higher correlations with the first two outcomes, in comparison to our results, could be due to the fact that Bayliss' study population was older, which resulted in a higher prevalence of chronic conditions. The floor effect in our "healthy" population might have attenuated the relation between the DBMA and these outcomes. We found a slightly higher correlation with depression than Bayliss and

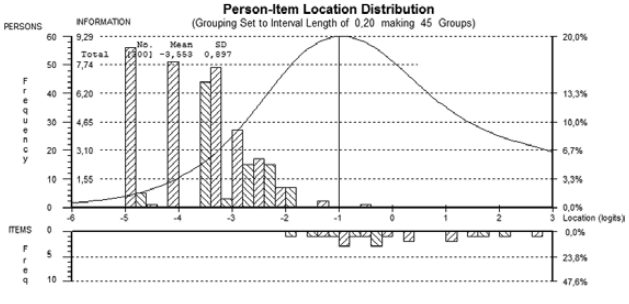


Figure 1. Person–item distribution and test information curve: final Rasch analysis of the Disease Burden Morbidity Assessment (DBMA).

Table 4. Relative Precision of the Linear Measure in Comparison to the Raw Summative DBMA Score (n = 1,277)

Scoring method	Patient groups	Mean (*SE*)	Mean difference (*SE*)	*Z* statistic	RP	95% CI
Raw score	<65 years	3.81 (0.20)	3.02 (0.35)	−9.89	1.00	
	≥65 years	6.83 (0.28)				
Linear measure	<65 years	5.95 (0.19)	2.98 (0.26)	−10.77	1.09	1.03–1.17
	≥65 years	8.92 (0.19)				
Raw score	Men	3.88 (0.19)	2.59 (0.35)	−6.61	1.00	
	Women	6.47 (0.28)				
Linear measure	Men	6.43 (0.19)	1.78 (0.27)	−6.33	0.96	0.86–1.05
	Women	8.21 (0.19)				

Note: CI = confidence interval; DBMA = Disease Burden Morbidity Assessment; RP = relative precision; *SE* = standard error.

Copyedited by: AM

colleagues, which can be ascribed to the fact that, unlike Bayliss and colleagues, we did include depression in the list of conditions used in the DBMA. Depression is a condition with a high prevalence and an important cause of disability (Ferrari et al., 2013; Moussavi et al., 2007). The inverse effect also exists: disability itself is an important predictor of depression (Bacon et al., 2016). Therefore, we argued that not including depression in the DBMA could underestimate disease burden scores. The relative precision analysis displayed some gain in precision in discriminating between age groups but when discriminating between sex groups, no difference was found with the original scale. These data suggest that the linear measure is at least as valid as the raw summative score concerning discriminant validity.

Some limitations must be acknowledged. We could not compare the outcomes of the DBMA with a "gold standard," because no other measures of multimorbidity or disease burden were included in the ELES-PS study. Secondly, as mentioned above, we validated a disease burden assessment instrument in a sample with quite a high health status, and little multimorbidity and disease burden. The studied population consisted of community-dwelling older adults, which means that institutionalized persons, with probably more multimorbidity, were not included. Also, only persons with household telephone lines were selected. The proportion of persons aged 50 years and older in Spain with telephone lines is estimated to be at least 92% (Rodríguez Laso et al., 2013), but it is possible that people that do not have a telephone line have lower incomes, which is known to be related to lower health status (Katz & Calasanti, 2015). Moreover, the persons that refused to participate in the CAPI interview, the questionnaire that contained the DBMA, were of higher age and reported lower perceived health than the respondents who did answer this questionnaire (Rodríguez Laso et al., 2013). Due to the relatively high health status, we found a very low PSI, high floor effects, and asymmetrical person–item threshold distribution. A third limitation was that, like other authors (Bayliss et al., 2009; Hudon, Fortin, Poitras, & Almirall, 2012; Poitras et al., 2012), we adapted the list of conditions included in the DBMA, which hinders comparisons with other studies. Further research should include the development of a standard list of conditions.

In summary, despite some limitations such as reliability below the expected and insufficient scale targeting, support was found for the validity of the DBMA as a patient-reported health outcome for measuring disease burden caused by 21 common chronic diseases in older adults. Its linear measure is related to patient-centered outcomes such as QoL and permits the calculation of change scores, making it potentially useful for the implementation of strategies to improve QoL and functional status among comorbid patients. Persons with multimorbidity are progressively becoming more common in our health care systems, and it is important to assess the impact that patients experience because of their multimorbidity. The DBMA measures the burden of multimorbidity, by asking the respondent to rate the impact of diseases on what is most important to patients themselves: their everyday life.

Supplementary Material

Supplementary data is available at *The Gerontologist* online.

Funding

This work was supported by the I+D+I National Plan, Spanish Ministry of Education and Science (CSO2011-30210-C02-01), the I+D+I National Plan of the Spanish Ministry of Economy, Industry and Competitiveness, co-funded by the European Regional Development Fund (ERDF) (ENVACES, ref: CSO2015-64115-R), and the I+D Activity Program of Madrid Community research groups on social sciences and humanities, co-funded by the European Social Fund (ENCAGE-CM, ref: S2015/HUM-3367).

Conflict of Interest

AQ5

The authors declare that there is no conflict of interest. Ethical approval: This study was approved by the Ethics Committee of the Spanish National Research Council.

References

AQ6

Bacon, K. L., Heeren, T., Keysor, J. J., Stuver, S. O., Cauley, J. A., & Fredman, L. (2016). Longitudinal and reciprocal relationships between depression and disability in older women caregivers and noncaregivers. *The Gerontologist*, **56**, 723–732. doi:10.1093/geront/gnu157

Barnett, K., Mercer, S. W., Norbury, M., Watt, G., Wyke, S., & Guthrie, B. (2012). Epidemiology of multimorbidity and implications for health care, research, and medical education: A cross-sectional study. *Lancet (London, England)*, **380**, 37–43. doi:10.1016/S0140-6736(12)60240-2

Bayliss, E. A., Ellis, J. L., & Steiner, J. F. (2005). Subjective assessments of comorbidity correlate with quality of life health outcomes: Initial validation of a comorbidity assessment instrument. *Health and Quality of Life Outcomes*, **3**, 51. doi:10.1186/1477-7525-3-51

Bayliss, E. A., Ellis, J. L., & Steiner, J. F. (2009). Seniors' self-reported multimorbidity captured biopsychosocial factors not incorporated into two other data-based morbidity measures. *Journal of Clinical Epidemiology*, **62**, 550–557. doi:10.1016/j.jclinepi.2008.05.002

Beard, J. R., Officer, A. M., & Cassels, A. K. (2016). The World Report on ageing and health. *The Gerontologist*, **56**(Suppl. 2), S163–S166. doi:10.1093/geront/gnw037

Bendayan, R., Cooper, R., Wloch, E. G., Hofer, S. M., Piccinin, A. M., & Muniz-Terrera, G. (2016). Hierarchy and speed of loss in physical functioning: A comparison across older U.S. and English men and women. *The Journals of Gerontology, Series A: Biological Sciences and Medical Sciences*, glw209. doi:10.1093/gerona/glw209

Copyedited by: AM

Byles, J. E., D'Este, C., Parkinson, L., O'Connell, R., & Treloar, C. (2005). Single index of multimorbidity did not predict multiple outcomes. *Journal of Clinical Epidemiology*, **58**, 997–1005. doi:10.1016/j.jclinepi.2005.02.025

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Routledge.

Crabtree, H. L., Gray, C. S., Hildreth, A. J., O'Connell, J. E., & Brown, J. (2000). The Comorbidity Symptom Scale: A combined disease inventory and assessment of symptom severity. *Journal of the American Geriatrics Society*, **48**, 1674–1678.

Ferrari, A. J., Charlson, F. J., Norman, R. E., Patten, S. B., Freedman, G., Murray, C. J., & Whiteford, H. A. (2013). Burden of depressive disorders by country, sex, age, and year: Findings from the global burden of disease study 2010. *PLoS Medicine*, **10**, e1001547. doi:10.1371/journal.pmed.1001547

Fitzpatrick, R., Norquist, J. M., Jenkinson, C., Reeves, B. C., Morris, R. W., Murray, D. W., & Gregg, P. J. (2004). A comparison of Rasch with Likert scoring to discriminate between patients' evaluations of total hip replacement surgery. *Quality of Life Research*, **13**, 331–338.

Forjaz, M. J., Ayala, A., Rodriguez-Blazquez, C., Prieto-Flores, M. E., Fernandez-Mayoralas, G., Rojo-Perez, F., & Martinez-Martin, P. (2012). Rasch analysis of the International Wellbeing Index in older adults. *International Psychogeriatrics*, **24**, 324–332. doi:10.1017/S104161021100158X

Forjaz, M. J., Ayala, A., Testa, C. M., Bain, P. G., Elble, R., Haubenberger, D., & Martinez-Martin, P. (2015). Proposing a Parkinson's disease–specific tremor scale from the MDS-UPDRS. *Movement Disorders*, **30**, 1139–1143. doi:10.1002/mds.26271

Fortin, M., Lapointe, L., Hudon, C., Vanasse, A., Ntetu, A. L., & Maltais, D. (2004). Multimorbidity and quality of life in primary care: A systematic review. *Health and Quality of Life Outcomes*, **2**, 51. doi:10.1186/1477-7525-2-51

Fortin, M., Stewart, M., Poitras, M. E., Almirall, J., & Maddocks, H. (2012). A systematic review of prevalence studies on multimorbidity: Toward a more uniform methodology. *Annals of Family Medicine*, **10**, 142–151. doi:10.1370/afm.1337

Fried, L. P., Bandeen-Roche, K., Kasper, J. D., & Guralnik, J. M. (1999). Association of comorbidity with disability in older women: The Women's Health and Aging Study. *Journal of Clinical Epidemiology*, **52**, 27–37.

Gadermann, A. M., Alonso, J., Vilagut, G., Zaslavsky, A. M., & Kessler, R. C. (2012). Comorbidity and disease burden in the National Comorbidity Survey Replication (NCS-R). *Depression and Anxiety*, **29**, 797–806. doi:10.1002/da.21924

Gijsen, R., Hoeymans, N., Schellevis, F. G., Ruwaard, D., Satariano, W. A., & van den Bos, G. A. (2001). Causes and consequences of comorbidity: A review. *Journal of Clinical Epidemiology*, **54**, 661–674.

Global Burden of Disease Study 2013 Collaborators. (2015). Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: A systematic analysis for the Global Burden of Disease Study 2013. *The Lancet*, **386**, 743–800. doi:10.1016/S0140-6736(15)60692-4

Groll, D. L., To, T., Bombardier, C., & Wright, J. G. (2005). The development of a comorbidity index with physical function as the outcome. *Journal of Clinical Epidemiology*, **58**, 595–602. doi:10.1016/j.jclinepi.2004.10.018

Hobart, J. C., Cano, S. J., Zajicek, J. P., & Thompson, A. J. (2007). Rating scales as outcome measures for clinical trials in neurology: Problems, solutions, and recommendations. *The Lancet. Neurology*, **6**, 1094–1105. doi:10.1016/S1474-4422(07)70290-9

Hudon, C., Fortin, M., Poitras, M. E., & Almirall, J. (2012). The relationship between literacy and multimorbidity in a primary care setting. *BMC Family Practice*, **13**, 33. doi:10.1186/1471-2296-13-33

Irwin, M., Artin, K. H., & Oxman, M. N. (1999). Screening for depression in the older adult: Criterion validity of the 10-item Center for Epidemiological Studies Depression Scale (CES-D). *Archives of Internal Medicine*, **159**, 1701–1704. doi:10.1001/archinte.159.15.1701

Katz, S., & Calasanti, T. (2015). Critical perspectives on successful aging: Does it "appeal more than it illuminates"? *The Gerontologist*, **55**, 26–33. doi:10.1093/geront/gnu027

Las Hayas, C., Bilbao, A., Quintana, J. M., Garcia, S., & Lafuente, I.; IRYSS Cataract Group. (2011). A comparison of standard scoring versus Rasch scoring of the visual function index-14 in patients with cataracts. *Investigative Ophthalmology & Visual Science*, **52**, 4800–4807. doi:10.1167/iovs.10-6132

Linacre, J. M. (1994). Sample size and item calibration stability. *Rasch Measurement Transactions*, **7**, 328.

Linacre, J. M. (2016). A user's guide to Winsteps Ministep Rasch-model computer programs. Retrieved from http://www.winsteps.com/manuals.htm

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, **47**, 149–174. doi:10.1007/BF02296272

Mavranezouli, I., Brazier, J. E., Young, T. A., & Barkham, M. (2011). Using Rasch analysis to form plausible health states amenable to valuation: The development of CORE-6D from a measure of common mental health problems (CORE-OM). *Quality of Life Research*, **20**, 321–333. doi:10.1007/s11136-010-9768-4

Moussavi, S., Chatterji, S., Verdes, E., Tandon, A., Patel, V., & Ustun, B. (2007). Depression, chronic diseases, and decrements in health: Results from the World Health Surveys. *Lancet (London, England)*, **370**, 851–858. doi:10.1016/S0140-6736(07)61415-9

Pallant, J. F., & Tennant, A. (2007). An introduction to the Rasch measurement model: An example using the Hospital Anxiety and Depression Scale (HADS). *The British Journal of Clinical Psychology*, **46**(Pt. 1), 1–18.

Pickard, A. S., De Leon, M. C., Kohlmann, T., Cella, D., & Rosenbloom, S. (2007). Psychometric comparison of the standard EQ-5D to a 5 level version in cancer patients. *Medical Care*, **45**, 259–263. doi:10.1097/01.mlr.0000254515.63841.81

Pinheiro, P., Plaß, D., & Krämer, A. (2011). The burden of disease approach for measuring population health. In A. Krämer, M. H. Khan, & F. Kraas (Eds.), *Health in Megacities and Urban Areas* (pp. 21–38). Heidelberg: Physica-Verlag.

Poitras, M. E., Fortin, M., Hudon, C., Haggerty, J., & Almirall, J. (2012). Validation of the disease burden morbidity assessment by self-report in a French-speaking population. *BMC Health Services Research*, **12**, 35. doi:10.1186/1472-6963-12-35

Rasch, G. (1980). *Probabilistic models for some intelligence and attainmant tests* (Expanded ed.). Chicago, IL: University of Chicago Press.

Robison, J., Gruman, C., Gaztambide, S., & Blank, K. (2002). Screening for depression in middle-aged and older Puerto Rican

Copyedited by: AM

primary care patients. *The Journals of Gerontology, Series A: Biological Sciences and Medical Sciences*, **57**, M308–M314.

Rodríguez Laso, Á., Urdaneta Artola, E., de la Fuente Sánchez, M., Galindo Moreno, E., Yanguas Lezáun, J. J., & Rodríguez Rodríguez, V. (2013). Análisis del sesgo de selección en el piloto de un estudio longitudinal sobre envejecimiento en España. *Gaceta Sanitaria*, **27**, 425–432. doi:10.1016/j.gaceta.2012.11.008

Sakthong, P., Charoenvisuthiwongs, R., & Shabunthom, R. (2008). A comparison of EQ-5D index scores using the UK, US, and Japan preference weights in a Thai sample with type 2 diabetes. *Health and Quality of Life Outcomes*, **6**, 71. doi:10.1186/1477-7525-6-71

Sangha, O., Stucki, G., Liang, M. H., Fossel, A. H., & Katz, J. N. (2003). The Self-Administered Comorbidity Questionnaire: A new method to assess comorbidity for clinical and health services research. *Arthritis and Rheumatism*, **49**, 156–163. doi:10.1002/art.10993

Smith, A. B., Rush, R., Fallowfield, L. J., Velikova, G., & Sharpe, M. (2008). Rasch fit statistics and sample size considerations for polytomous data. *BMC Medical Research Methodology*, **8**, 33. doi:10.1186/1471-2288-8-33

Smith, A. B., Wright, E. P., Rush, R., Stark, D. P., Velikova, G., & Selby, P. J. (2006). Rasch analysis of the dimensional structure of the Hospital Anxiety and Depression Scale. *Psycho-Oncology*, **15**, 817–827. doi:10.1002/pon.1015

Tennant, A., & Conaghan, P. G. (2007). The Rasch measurement model in rheumatology: What is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis and Rheumatism*, **57**, 1358–1362. doi:10.1002/art.23108

Tennant, A., McKenna, S. P., & Hagell, P. (2004). Application of Rasch analysis in the development and application of quality of life instruments. *Value in Health*, **7**, S22–S26. doi:10.1111/j.1524-4733.2004.7s106.x

Tennant, A., & Pallant, J. F. (2007). DIF matters: A practical approach to test if differential item functioning makes a difference. *Rasch Measurement Transactions*, **20**, 1082–1084.

Tennant, A., Penta, M., Tesio, L., Grimby, G., Thonnard, J. L., Slade, A., & Phillips, S. (2004). Assessing and adjusting for cross-cultural validity of impairment and activity limitation scales through differential item functioning within the framework of the Rasch model: The PRO-ESOR project. *Medical Care*, **42**(1 Suppl.), I37–I48. doi:10.1097/01.mlr.0000103529.63132.77

Teófilo Rodríguez, J., González Cabezas, A. N., Díaz Veiga, P., & Rodríguez Rodríguez, V. (2011). Estudio Longitudinal Envejecer en España: El proyecto ELES. *Perfiles y Tendencias*, **50**, 1–44.

The International Wellbeing Group. (2013). *Personal Wellbeing Index-Adult (PWI-A). Manual 2013* (5th ed.). Melbourne: Australian Centre on Quality of Life, Deakin University. Retrieved from http://www.deakin.edu.au/research/acqol/instruments/wellbeing-index/pwi-a-english.pdf

United Nations, Department of Economic and Social Affairs, Population Division. (2013). *World Population Ageing 2013*. Retrieved from http://www.un.org/en/development/desa/population/publications/pdf/ageing/WorldPopulationAgeing2013.pdf

Virués-Ortega, J., Rodríguez-Blázquez, C., Micheli, F., Carod-Artal, F. J., Serrano-Dueñas, M., & Martínez-Martín, P. (2010). Cross-cultural evaluation of the modified Parkinson Psychosis Rating Scale across disease stages. *Movement Disorders*, **25**, 1391–1398. doi:10.1002/mds.23081

Wijers, I. G., Ayala, A., Rodriguez-Blazquez, C., Rodriguez-Laso, A., Rodriguez-Rodriguez, V., & Forjaz, M. J. (2016). Disease burden morbidity assessment by self-report: Psychometric properties in older adults in Spain. *Geriatrics & Gerontology International*. Advanced online publication. doi:10.1111/ggi.12835